第一章　线性空间中的凸集

§1.1　线性空间及其代数对偶

线性空间是通常空间的推广．它把通常的向量加法与放大、缩小(数乘)抽象化，形成线性空间的公理．下面在一般的线性空间中讨论凸集，其有关结果当然首先在通常的空间中成立．

一般的线性空间可在任意域上定义．本书只讨论实数域$\mathbf{R}$上的线性空间，以后我们提到的线性空间都是指实线性空间．

我们称 X 为(实)线性空间，是指 X 为满足下列性质的集合：

(I)　X 上有加法运算$+$，即

$\forall x,\ y\in X,\ \exists z=x+y\in X$，且满足：

1)　$\forall x,\ y\in X,\ x+y=y+x$；

2)　$\forall x,\ y,\ z\in X,\ x+(y+z)=(x+y)+z$；

3)　$\exists 0\in X,\ \forall x\in X,\ x+0=x$；

4)　$\forall x\in X,\ \exists -x\in X,\ x+(-x)=x-x=0$．

(II)　X 上有(实)数乘运算，即

$\forall x\in X,\ \forall\lambda\in\mathbf{R},\ \exists y=\lambda x\in X$，且满足

1)　$\forall\lambda,\ \mu\in\mathbf{R},\ \forall x\in X,\ \lambda(\mu x)=(\lambda\mu)x$；

2)　$\forall x\in X,\ 1\cdot x=x$．

(III)　加法运算和数乘运算满足分配律，即

1)　$\forall\lambda,\ \mu\in\mathbf{R},\ \forall x\in X,\ (\lambda+\mu)x=\lambda x+\mu x$；

2)　$\forall\lambda\in\mathbf{R},\ \forall x,\ y\in X,\ \lambda(x+y)=\lambda x+\lambda y$．

X 中的元素称为 X 的点或向量．记号 0 既表示实数零，也表示(I)3)中的零向量．X 上的加法和数乘运算则称为线性运算．

如果对于 $x_1,\ \cdots,\ x_n\in X$，存在不全为零的 $\lambda_1,\ \cdots,\ \lambda_n\in\mathbf{R}$，使得

$$\lambda_1x_1+\cdots+\lambda_nx_n=0,$$

则称这 n 个向量线性相关；如果这样的实数组不存在，则称这 n 个向量线性无关．如果 X 的子集 $A\subset X$ 中任何有限个向量都线性无关，则称 A 为 X 的线性无关集（又称代数自由集）．如果线性空间 X 中不存在元素个数多于 n 个的线性无关集，但存在 n 个线性无关的向量，则称空间 X 是 n 维线性空间．如果这样的自然数 n 不存在，则称空间 X 为无限维线性空间．

在一般的数学规划问题中，通常只涉及有限维线性空间；而在变分学、最优控制、逼近论等学科中的极值问题，通常在无限维线性空间中讨论．不过，空间的维数在很多情形对问题的讨论并无影响．因此，若在一般的线性空间框架中讨论，就能得到一些对规划论、变分学、最优控制、逼近论等都适用的结果．

由于我们希望能得到对任何维数的线性空间都能成立的结果，势必需要用到有关选择公理的推理．这时，我们总采用与选择公理等价的下列引理：

Zorn 引理 如果序集 M 中每一全序子集都有上界，那么 M 有极大元．▌

这里非空集 M 称为序集，是指在 M 上定义了一个序关系 $\leqslant$，满足：

1) $\forall x\in M,\ x\leqslant x$；

2) $\forall x,\ y\in M,\ \{x\leqslant y,\ y\leqslant x\}\Rightarrow\{x=y\}$；

3) $\forall x,\ y,\ z\in M,\ \{x\leqslant y,\ y\leqslant z\}\Rightarrow\{x\leqslant z\}$；

如果对于序集 M，还满足

4) $\forall x,\ y\in M$，或者 $x\leqslant y$，或者 $y\leqslant x$．

那么 M 就称为全序集．序集 M 的全序子集 N，是指 $N\subset M$ 关于 M 中的序关系构成全序集．$x_0\in M$ 称为 N 的上界，是指 $\forall x\in N$，$x\leqslant x_0$．$x_0\in M$ 称为 M 的极大元，是指不存在 $x_1\in M$，$x_1\neq x_0$，使得 $x_0\leqslant x_1$．

下面是运用 Zorn 引理的一个例子．如果只需要得到有关有限维空间的结论，那么只要运用数学归纳法就能得到类似的结果．在"无限推理"中，Zorn 引理的作用与数学归纳法的作用相似．

定理 1.1.1 任何维数不为零的线性空间 X 有 Hamel 基，即存在线性无关集 $\{e_i\}_{i\in I}\subset X$，使得

$\forall x\in X$，$\exists x^i\in\mathbf{R}$，$i\in I$，只有有限个不为零，

$$x=\sum_{i\in I}x^ie_i.$$

式中 $\sum$ 对有限个非零元素求和. 此外，还可要求 $\{e_i\}_{i\in I}$ 包含给定的线性无关集 $E\subset X$ 作为其子集.

证明 设 M 为 X 中所有包含 E 的线性无关集全体，并在 M 中以包含关系作为序关系，即

$$\forall A、B\in M,\ A\leqslant B\Leftrightarrow A\subset B,$$

则 M 成为一个序集，M 的每一全序子集都有上界. 事实上，设 $\{A_j\}_{j\in J}\subset M$ 为 M 的全序子集，则 A_j，$j\in J$ 都是 X 中的包含 E 的线性无关集，且

$$\forall j_1,\ j_2\in J,\ \text{或者}\ A_{j_1}\subset A_{j_2},\ \text{或者}\ A_{j_2}\subset A_{j_1}.$$

于是 $\bigcup_{j\in J}A_j$ 显然是 $\{A_j\}_{j\in J}$ 的上界. 因此，由 Zorn 引理，M 中有极大元. 设

$$B=\{e_i\}_{i\in I}\in M$$

为 M 的极大元. 那么 $E\subset B$，且 B 必定是 X 中的 Hamel 基. 否则，存在 $y\in X$，不能用 B 中的元素来线性表示，即 y 与 B 中所有元素组都线性无关. 这样，$\{y\}\cup B\in M$，且 $B\subset\{y\}\cup B$，与 B 的极大性矛盾. ∎

当 X 为 n 维空间时，它的 Hamel 基中的向量个数显然为 n 个. 设这 n 个基向量为 e_1，…，e_n. 则任何 $x\in X$ 都可唯一表示为

$$x=x^1e_1+\cdots+x^ne_n,\quad \text{其中}\ x^1,\ \cdots,\ x^n\in\mathbf{R},$$

这里唯一性可由 e_1，…，e_n 线性无关而得到. 这样，每个 $x\in X$ 都可与实数组 $(x^1,\ \cdots,\ x^n)$ 一一对应，而且这一对应保持线性运算关系. 实数组 $(x^1,\ \cdots,\ x^n)$ 全体可看作 n 个实数域 $\mathbf{R}$ 的乘积集合 $\mathbf{R}^n$. 因此，每个 n 维空间都可在"同构意义"下看作 $\mathbf{R}^n$. 以后，我们就将总是以 $\mathbf{R}^n$ 表示 n 维线性空间.

设 X、Y 为两个线性空间. 映射 A: $X\to Y$ 称为线性映射，

是指

$$\forall \lambda, \mu \in \mathbf{R}, \forall x, y \in X,$$
$$A(\lambda x+\mu y)=\lambda A(x)+\mu A(y).$$

当 X、Y 分别为有限维空间 $\mathbf{R}^n$、$\mathbf{R}^m$ 时，利用 X 中的基 $\{e_i\}_{i=1,\cdots,n}$，和 Y 中的基 $\{f_j\}_{j=1,\cdots,m}$，设

$$Ae_i=a_{i1}f_1+\cdots+a_{im}f_m \quad (i=1, \cdots, n).$$

则线性映射 A 完全可由矩阵 $\{a_{ij}\}_{\substack{i=1,\cdots,n\\ j=1,\cdots,m}}$ 来刻划．对于一般的有 Hamel 基 $\{e_i\}_{i\in I}$ 的 X 和有 Hamel 基 $\{f_j\}_{j\in J}$ 的 Y，线性映射 $A: X\to Y$ 也完全可由无限矩阵 $\{a_{ij}\}_{\substack{i\in I\\ j\in J}}$ 来刻划．这里对于固定的 i，只有有限个 a_{ij} 不为零．

有一类特殊的线性映射特别重要，这就是 $Y=\mathbf{R}$ 的情形．这种线性映射称为 X 上的线性形式（又称线性泛函）．它将表示为 $x\mapsto\langle x^*, x\rangle\in\mathbf{R}$ 的形式，即 $\langle x^*, x\rangle$ 满足

$$\forall \lambda, \mu \in \mathbf{R}, \forall x, y \in X,$$
$$\langle x^*, \lambda x+\mu y\rangle=\lambda\langle x^*, x\rangle+\mu\langle x^*, y\rangle.$$

X 上的线性形式 x^* 的全体记作 X^*，它称为 X 的代数对偶空间．如果在 X^* 中定义加法和数乘如下：

$$\forall \lambda、\mu \in \mathbf{R}, \forall x^*、y^* \in X^*, \forall x \in X;$$
$$\langle\lambda x^*+\mu y^*, x\rangle=\lambda\langle x^*, x\rangle+\mu\langle y^*, x\rangle,$$

那么 X^* 也是线性空间．

如上所述，设 X 的 Hamel 基为 $\{e_i\}_{i\in I}$，那么每一 $x^*\in X^*$ 都可用数组 $\{x_i^*\}_{i\in I}$，$x_i^*\in\mathbf{R}$ 来刻划，即如果

$$x=\sum_{i\in I} x^i e_i,$$

这里 x^i 中只有有限个不为零，那么

$$\langle x^*, x\rangle=\sum_{i\in I} x_i^*\cdot x^i.$$

特别是当 $X=\mathbf{R}^n$ 时，所有 $x^*\in\mathbf{R}^{n*}$ 都可用

$$\{x_i^*\}_{i=1,\cdots,n}=(x_1^*, \cdots, x_n^*)$$

来表示．不难看出，$\mathbf{R}^{n*}$ 也是 n 维线性空间，或者说，$\mathbf{R}^n$ 和 $\mathbf{R}^{n*}$ 是“同构”的，而 $\langle\cdot, \cdot\rangle$ 可理解为 n 维空间上的内积．但是以后我们

还是经常把 $\mathbf{R}^n$ 和 $\mathbf{R}^{n*}$ 看作不同的空间．当 X 为无限维空间时，X^* 与 X 就很不一样．事实上，这时 X 的向量 x 的表示 $\{x^i\}_{i\in I}$ 中只有有限个 x^i 不为零，而 X^* 的向量 x^* 的表示 $\{x_i^*\}_{i\in I}$ 中可以有无限个 x_i^* 不为零．由此可见，这时 X^* 要比 X 复杂[*)]．

对于任何 $x\in X$，$x^*\mapsto\langle x^*, x\rangle$ 也可看作 X^* 上的线性形式；也就是说，设 X^* 的代数对偶空间为 X^{**}，那么可以认为 $X\subset X^{**}$．下列命题指出了无限维空间的一个特征：

命题 1.1.2 线性空间 X 为无限维的当且仅当 $X\neq X^{**}$．

证明 设 X 有无限 Hamel 基为 $\{e_i\}_{i\in I}$．令

$$\{e^{*j}\}_{j\in I}\subset X^*,$$

满足

$$\langle e^{*j}, e_i\rangle=\begin{cases}1, & \text{当 } i=j\\ 0, & \text{当 } i\neq j\end{cases}, \quad \forall i, j\in I. \tag{1.1.1}$$

则不难验证，所有 e^{*j} 之间都线性无关．令 $\{e^{*j'}\}_{j'\in J}$ 为包含 $\{e^{*j}\}_{j\in I}$ 的 X^* 的 Hamel 基．那么任何 $x^{**}\in X^{**}$ 都可由实数组

$$x^{**j'}=\langle x^{**}, e^{*j'}\rangle, \quad \forall j'\in J$$

来刻划，这里 $x^{**j'}$ 可取任何实数．取 $\hat{x}^{**}\in X^{**}$ 满足

$$\hat{x}^{**j'}=\langle \hat{x}^{**}, e^{*j'}\rangle=1, \quad \forall j'\in J.$$

那么 $\hat{x}^{**}$ 不可能是 X 中的元素，因为每一 $x\in X$，都有表示式

$$x=\sum_{i\in I}x^i e_i,$$

这里 x^i 中只有有限个不为零；由 e^{*j} 的定义(1.1.1)，$\langle e^{*j}, x\rangle$ 中也只能有有限个不为零．因此，$X\neq X^{**}$．

另一方面，当 X 为有限维时，必定有 $X=X^{**}$．故命题得证． ∎

最后，我们再提出两个线性空间 X、Y 的乘积空间 $X\times Y$ 的概念．$X\times Y$ 定义为

$$X\times Y=\{(x, y)\mid x\in X, y\in Y\},$$

其上的加法和数乘定义为

---

*) 更确切地说，X^* 的 Hamel 基的势大于 X 的 Hamel 基的势．可以证明，当 X 为无限维时，X^* 的 Hamel 基的势严格大于 X 的 Hamel 基的势．

$$\forall x_1, x_2 \in X, \forall y_1, y_2 \in Y, \forall \lambda, \mu \in \mathbf{R},$$
$$\lambda(x_1, y_1) + \mu(x_2, y_2) = (\lambda x_1 + \mu x_2, \lambda y_1 + \mu y_2).$$

这样，$X \times Y$ 也是线性空间.

设 $(e_i^X)_{i\in I}$，$(e_j^Y)_{j\in J}$ 分别为 X、Y 的 Hamel 基. 则不难验证

$$\{(e_i^X, 0)\}_{i\in I} \cup \{(0, e_j^Y)\}_{j\in J}$$

构成 $X\times Y$ 的 Hamel 基. 从而任何 $z^* \in (X\times Y)^*$ 可由下列实数组

$$\langle z^*, (e_i^X, 0)\rangle, \forall i\in I; \langle z^*, (0, e_j^Y)\rangle, \forall j\in J$$

来刻划. 实数组 $\langle z^*, (e_i^X, 0)\rangle$, $i\in I$ 和 $\langle z^*, (0, e_j^Y)\rangle$, $j\in J$ 分别可与 X^* 和 Y^* 中的元素相对应. 这就可看出下列命题成立:

命题 1.1.3 设 X、Y 为两个线性空间，$X\times Y$ 为它们的乘积线性空间. 那么 $(X\times Y)^*$ 可看作 X^* 与 Y^* 的乘积空间 $X^*\times Y^*$，即 $(X\times Y)^*$ 中的每一元素都与某个 $(x^*, y^*) \in X^*\times Y^*$ 一一对应，且

$$\langle (x^*, y^*), (x, y)\rangle = \langle x^*, x\rangle + \langle y^*, y\rangle,$$
$$\forall x\in X, \forall y \in Y. \blacksquare$$

以后我们经常要利用 $Y=\mathbf{R}$ 时的命题 1.1.3; 这时 $\langle y^*, y\rangle$ 就是两个实数相乘.

§1.2 子空间、仿射集、锥和凸集

设 X 为线性空间，$L\subset X$. 如果 L 对 X 的线性运算也构成线性空间，那么 L 称为 X 的(线性)子空间. L 为 X 的子空间显然当且仅当

$$\forall x_1, x_2 \in L, \forall \lambda_1, \lambda_2 \in \mathbf{R}, \lambda_1 x_1 + \lambda_2 x_2 \in L. \qquad (1.2.1)$$

引入下列记号:

$$A\pm B = \{x\in X \mid x = x_1 \pm x_2, x_1\in A, x_2\in B\};$$
$$\lambda A = \{x\in X \mid x = \lambda x_1, x_1 \in A\};$$
$$\forall A, B\subset X, \forall \lambda\in \mathbf{R}.$$

那么条件(1.2.1)也可写成

$$\lambda_1 L+\lambda_2 L\subset L,\ \forall\lambda_1,\ \lambda_2\in\mathbf{R}. \qquad (1.2.2)$$

很明显，它也等价于对于任何自然数 $n\geqslant 2$,

$$\lambda_1 L+\lambda_2 L+\cdots+\lambda_n L\subset L,\ \forall\lambda_1,\ \cdots,\ \lambda_n\in\mathbf{R}. \qquad (1.2.2)'$$

X 上的每一线性形式 $x^*\in X^*$ 在 L 上的限制都可看作 L 的线性形式. 反之，每一 L 上的线性形式 $x_L^*\in L^*$ 也一定可扩充为 X^* 的元素，甚至还可有一定的特殊要求，从而有

命题 1.2.1 设 L 为线性空间 X 的真子空间，即

$$L\neq X,\ x_L^*\in L^*$$

为 L 上的线性形式. 则对于任何 $x_0\in X\setminus L$, 存在 $x^*\in X^*$, 使得

$$\langle x^*,\ x_0\rangle=1;\ \ \langle x^*,\ x\rangle=\langle x_L^*,\ x\rangle,\ \forall x\in L.$$

证明 设 L 的 Hamel 基为 $\{e_i^L\}_{i\in I}$. 由 $x_0\in X\setminus L$, 可知

$$\{x_0\}\cup\{e_i^L\}_{i\in I}$$

为线性无关集. 根据定理 1.1.1, 它可扩充为 X 的 Hamel 基 $\{x_0\}\cup\{e_j\}_{j\in J}$, 这里有 $\{e_i^L\}_{i\in I}\subset\{e_j\}_{j\in J}$. 任何 $x^*\in X^*$ 都可由实数组 $\{\langle x^*,\ x_0\rangle\}\cup\{\langle x^*,\ e_j\rangle\}_{j\in J}$ 唯一确定. 于是任何满足

$$\langle x^*,\ x_0\rangle=1;\ \ \langle x^*,\ e_i^L\rangle=\langle x_L^*,\ e_i^L\rangle,\ \forall i\in I$$

的 x^* 都满足命题要求. ∎

X 的子空间的平移称为 X 的**仿射集**(又称**线性流形**), 即 $A\subset X$ 为仿射集，是指

$$A=\{x_0\}+L=x_0+L.$$

这里 L 为 X 的子空间，$x_0\in X$ (右端{ }被省略). L 的维数也称为 A 的维数.

设 x_1、$x_2\in X$. 我们称

$$\overline{x_1x_2}=\{x\in X\,|\,x=x_1+\lambda(x_2-x_1),\ \lambda\in\mathbf{R}\} \qquad (1.2.3)$$

为连结 x_1、x_2 的直线. 于是，仿射集有下列特征:

命题 1.2.2 $A\subset X$ 为线性空间 X 的仿射集当且仅当

$$\forall x_1,\ x_2\in A,\ \overline{x_1x_2}\subset A. \qquad (1.2.4)$$

证明 设 $A=x_0+L$, 这里 $x_0\in X$, L 为 X 的子空间. 那么 $L=A-x_0$, 从而对于任何 x_1、$x_2\in A$, $\lambda\in\mathbf{R}$, 有

$$(1-\lambda)(x_1-x_0)+\lambda(x_2-x_0)\in L=A-x_0,$$

即存在 $\hat{x}\in A$，使得

$$(1-\lambda)x_1+\lambda x_2-x_0=\hat{x}-x_0.$$

从而 $\quad (1-\lambda)x_1+\lambda x_2\in A,\ \forall\lambda\in\mathbf{R},\ \forall x_1,\ x_2\in A,$

即(1.2.4)成立.

反之，如果(1.2.4)成立，那么任取 $x_0\in A$，对于任何 $\lambda\in\mathbf{R}$，$x_1\in A$，有

$$(1-\lambda)x_0+\lambda x_1\in A;$$

对于任何 $\lambda_2\in\mathbf{R}$，$x_2\in A$，有

$$(1-\lambda_2)[(1-\lambda)x_0+\lambda x_1]+\lambda_2x_2\in A.$$

令 $\lambda(1-\lambda_2)=\lambda_1$，则由上式可得

$$(1-\lambda_1-\lambda_2)x_0+\lambda_1x_1+\lambda_2x_2\in A,$$

$$\forall\lambda_1,\ \lambda_2\in\mathbf{R},\ \forall x_1,\ x_2\in A;$$

或 $$\lambda_1(x_1-x_0)+\lambda_2(x_2-x_0)\in A-x_0,$$

$$\forall\lambda_1,\ \lambda_2\in\mathbf{R},\ \forall x_1\ x_2\in A.$$

由(1.2.1)，$L=A-x_0$ 是线性子空间. ▮

注 在证明的后半部分中，x_0 是任意的. 因此，我们有

$$L=A-A.$$

这样，由(1.2.3)和(1.2.4)，A 为仿射集的充要条件也可记为

$$(1-\lambda)A+\lambda A\subset A,\ \forall\lambda\in\mathbf{R}. \tag{1.2.5}$$

如同命题1.2.2的证明中那样，重复利用参数 λ，可得(1.2.5)等价于对于任何 $n\geqslant2$，

$$\lambda_1A+\lambda_2A+\cdots+\lambda_nA\subset A,\forall\lambda_1,\ \cdots,\ \lambda_n\in\mathbf{R},\ \sum_{i=1}^{n}\lambda_i=1. \tag{1.2.5$'$}$$

一类重要的仿射集有下列形式：

$$H=\{x\in X\mid\langle x^*,\ x\rangle=\alpha\},$$

这里 $x^*\in X^*$，$x^*\neq0$，$\alpha\in\mathbf{R}$. 它们都是子空间

$$H_0=\{x\in X\mid\langle x^*,\ x\rangle=0\}$$

的平移. 这类仿射集称为 X 的(由 $(x^*,\ \alpha)$ 决定的)**超平面**，它们是通常平面的一般化.

在线性空间中，除直线(1.2.3)外，我们还能定义射线和线段。设 x_1、$x_2\in X$。我们称

$$\overrightarrow{x_1x_2}=\{x\in X\mid x=x_1+\lambda(x_2-x_1),\ \lambda>0\}$$

为连结 x_1、x_2 的以 x_1 为顶点的射线；称

$$[x_1,\ x_2]=\{x\in X\mid x=x_1+\lambda(x_2-x_1),\ \lambda\in[0,\ 1]\}$$

为连结 x_1、x_2 的线段或闭区间。类似地还可定义开区间、半开半闭区间：

$$(x_1,\ x_2)=\{x\in X\mid x=x_1+\lambda(x_2-x_1),\ \lambda\in(0,\ 1)\};$$

$$[x_1,\ x_2)=\{x\in X\mid x=x_1+\lambda(x_2-x_1),\ \lambda\in[0,\ 1)\},$$

等等。这些概念的几何意义都是明显的。

如果集合 $C\subset X$，满足

$$\forall x\in C,\quad \overrightarrow{x_1x}\subset C, \tag{1.2.6}$$

那么 C 称为 X 的以 x_1 为顶点的锥。特别是，当 $x_1=0$ 时，就简称为锥。这时，(1.2.6)变为

$$\forall x\in C,\quad \overrightarrow{0x}\in C,$$

或

$$\forall x\in C,\ \forall\lambda>0,\ \lambda x\in C. \tag{1.2.7}$$

如果集合 K 满足

$$\forall x_1、x_2\in K,\ [x_1,\ x_2]\subset K, \tag{1.2.8}$$

那么 K 称为 X 的凸集。同时为凸集的锥则称为凸锥。由线段的定义，(1.2.8)也可写成

$$(1-\lambda)K+\lambda K\subset K,\ \forall\lambda\in[0,\ 1]. \tag{1.2.9}$$

而由(1.2.7)和(1.2.9)可知，集合 C 为凸锥的充要条件为

$$\lambda_1C+\lambda_2C\subset C,\ \forall\lambda_1,\ \lambda_2>0. \tag{1.2.10}$$

把(1.2.9)和(1.2.10)迭代若干次，我们就可把线性子空间、仿射集、凸集和凸锥的条件写成非常相似的形式，即对于任何 $n\geqslant2$，有

线性子空间 L：

$$\lambda_1L+\cdots+\lambda_nL\subset L,\ \forall\lambda_1,\ \cdots,\ \lambda_n\in\mathbf{R};$$

仿射集 A：

$$\lambda_1 A+\cdots+\lambda_m A\subset A,\ \forall\lambda_1,\ \cdots,\ \lambda_n\in\mathbf{R},\ \sum_{i=1}^{n}\lambda_i=1;$$

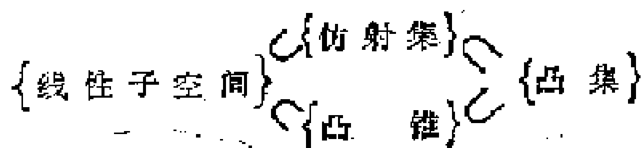
凸集 K:

$$\lambda_1 K+\cdots+\lambda_m K\subset K,\ \forall\lambda_1,\ \cdots,\ \lambda_n\in[0,\ 1],\ \sum_{i=1}^{n}\lambda_i=1;$$

凸锥 C:

$$\lambda_1 C+\cdots+\lambda_n C\subset C,\ \forall\lambda_1,\ \cdots,\ \lambda_n>0.$$

由此也立即可看出:



{仿射集}∩{凸锥}={线性子空间}

注 上面四个条件中的 ⊂，除了线性子空间外，都能换为 =。对于线性子空间来说，当所有 $\lambda_i=0$ 时，等号就不成立；但除此以外，⊂ 也可换为等号。此外，对于凸集和凸锥来说，其中可能存在端元素，即不能表示为其他元素的"凸组合"或"凸锥组合"的元素(例如，$\mathbf{R}^1$ 中的闭区间的两个端点；$\mathbf{R}^2$ 中的扇形的两条边)，但它们总能表示为自身的"凸组合"或"凸锥组合"。▌

下列命题由定义立即可得:

命题 1.2.3 设 $A_j\in X(j=1,\ \cdots,\ m)$ 为 X 的线性子空间(仿射集、锥、凸集)，$\mu_j\in\mathbf{R}(j=1,\ \cdots,\ m)$。那么 $\mu_1A_1+\cdots+\mu_mA_m$ 也是线性子空间(仿射集、锥、凸集)。▌

命题 1.2.4 设 $A_j\in X,\ j\in J$ 为 X 的线性子空间(仿射集、锥、凸集)，这里 J 为任意指标集。那么，$\bigcup_{j\in J}A_j$ 也是线性子空间(仿射集、锥、凸集)。▌

注 对于集合的并，命题 1.2.4 不成立。▌

设 $E\subset X$ 为 X 的任意子集。我们称包含 E 的最小线性子空间(仿射集、锥、凸集)为 E 的线性包(仿射包、锥包、凸包)，记作 $\mathrm{lin}E$ ($\mathrm{aff}E$, $\mathrm{cone}E$, $\mathrm{co}E$)。设 $A_j,\ j\in J$ 为所有包含 E 的线性子

空间(仿射集、锥、凸集). 那么 E 的线性包(仿射包、锥包、凸包)就是 $\bigcap_{j\in J} A_j$. 它们也可直接表示如下:

$$\operatorname{lin} E=\{x\in X \mid x=\sum_{y\in E}\lambda_y y,\ \lambda_y\in\mathbb{R},\ \text{且仅有有限个不为零}\}$$

$$\operatorname{aff} E=\{x\in X \mid x=\sum_{y\in E}\lambda_y y,\ \lambda_y\in\mathbb{R},$$

$$\text{仅有有限个不为零, 且} \sum_{y\in E}\lambda_y=1.\};$$

$$\operatorname{cone} E=\{x\in X \mid x=\lambda y,\ y\in E,\ \lambda>0\}=\bigcup_{\lambda>0}\lambda E;$$

$$\operatorname{co} E=\{x\in X \mid x=\sum_{y\in E}\lambda_y y,\ \lambda_y\in[0,1],$$

$$\text{仅有有限个不为零, 且} \sum_{y\in E}\lambda_y=1.\}.$$

有时也称它们为**由 E 张成的线性子空间、仿射集、锥或凸集**.

有限个点张成的凸集称为**凸多面体**. 其重要的特殊情形是 n 维**单纯形**, 它是由 $n+1$ 个"仿射无关"(即不在同一个 $n-1$ 维仿射集上)的点张成的凸多面体. 例如, 一维单纯形是线段; 二维单纯形是三角形; 三维单纯形是四面体等.

§1.3 代数内点、代数开集和代数闭集

如上所述, 在线性空间中有直线的概念, 如开区间、闭区间等, 它们也能直接给线性空间带来某些拓扑概念. 但这样引进的拓扑概念有时会与真正的拓扑概念不一致(参看 §3.1). 因此, 有关的名称我们都冠以"代数"二字.

设 $A\subset X$ 为线性空间 X 的集合. 如果点 $a\in A$, 满足

$$\forall h\in X,\ \exists\varepsilon>0,\ (a-\varepsilon h,\ a+\varepsilon h)\subset A, \tag{1.3.1}$$

那么称 a 为 A 的**代数内点**. A 的代数内点全体称为 A 的**代数内部**[又称**核**(core)], 记作 A^i. 如果 $A=A^i$, 那么 A 称为 X 的**代数开集**. 条件(1.3.1)也显然等价于:

$$\forall h\in X,\ \exists\varepsilon>0,\ [a,\ a+\varepsilon h]\subset A. \tag{1.3.2}$$

在证明有关代数内点的结果时, 后者更方便些. 而代数内部 A^i 也

可直接定义为：

$$A^i=\{x\in X\mid \forall h\in X,\ \exists\varepsilon>0,\ \forall t\in[0,\ \varepsilon],\ x+th\in A\}. \tag{1.3.3}$$

由于代数内点的概念只涉及过这点的直线，因此可把这个概念推广到仿射集上，或者说，推广到 A 的仿射包 aff A 上.

我们称 $a\in A$ 为 A 的相对代数内点，是指

$$\forall h\in \text{aff } A-a(=\text{aff } A-A),\ \exists\varepsilon>0,$$

$$(a-\varepsilon h,\ a+\varepsilon h)\subset A \quad 或 \quad [a,\ a+\varepsilon h]\subset A.$$

A 的相对代数内点的全体称为 A 的相对代数内部，记作 A^{ri}. 这样，例如平面上的一条直线对平面来说，没有代数内点，但其所有点都是相对代数内点. 还需注意，任何单点集的相对内部是非空的.

命题 1.3.1 i) 如果 $A_1,\ \cdots,\ A_n$ 为代数开集，那么 $\bigcap\limits_{k=1}^{n} A_k$ 也是代数开集；

ii) 如果 $A_k,\ k\in I$ 为代数开集，那么 $\bigcup\limits_{k\in I} A_k$ 也是代数开集.

证明直接由定义可得. ▌

注 1 根据拓扑空间的定义(参看 § 3.1)，如果再规定空集 $\varnothing$ 也是代数开集，那么命题 1.3.1 说明，线性空间 X 对它的代数开集族形成拓扑空间. 但这是一种非常特殊的拓扑. 尤其是对于有限维空间来说，它比通常的由 Euclid 距离决定的拓扑强得多，即通常理解的开集一定是代数开集，但反之不然. 例如，在 $\mathbf{R}^2$ 中，令

$$A=\Big\{x=(x^1,\ x^2)\in\mathbf{R}^2\mid (x^1)^2+(x^2)^2<1,$$
$$\Big(x^1-\frac{1}{2}\Big)^2+(x^2)^2\neq\frac{1}{4}\Big\}\cup\{0\},$$

(如图 1.1). 那么 A 是代数开集，尤其是 $0\in A$ 是 A 的代数内点. 但它不是通常的开集. ▌

注 2 对于一个拓扑空间的子集 A 来说，A 的(拓扑)内部 int A 是指包含在 A 内的最大开集. 不言而喻，int A 总是开集，容易使人发生错觉，以为“线性空间 X 的集合 A 的代数内部 A^i 就

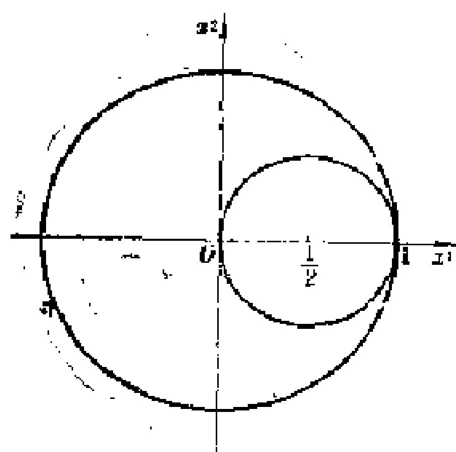


图 1.1

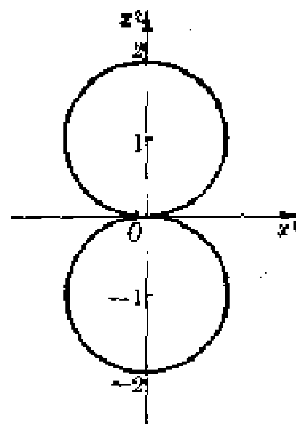


图 1.2

是包含在 A 内的最大代数开集". 这是不正确的，因为 A^i 一般不一定是代数开集. 下面就是一个反例.

例 $X=\mathbf{R}^2$.

$$A=\{x\in\mathbf{R}^2\mid(x^1)^2+(x^2-1)^2\leqslant 1\}\cup\{x\in\mathbf{R}^2\mid(x^1)^2+(x^2+1)^2\leqslant 1\}\cup\{x\in\mathbf{R}^2\mid x^2=0\}(\text{如图 } 1.2),$$

$$A^i=\{x\mid(x^1)^2+(x^2-1)^2<1\}\cup\{x\mid(x^1)^2+(x^2+1)^2<1\}\cup\{0\},$$

$$(A^i)^i=A^i\setminus\{0\}\neq A^i.$$

这个例子说明代数内部确实是个"代数"概念，而不是拓扑概念. 因而我们在线性空间框架中讨论的结果一般不能看作是在拓扑线性空间(参看§3.1)框架中的结果的特例. ▮

注 3 造成上面这种怪现象的主要原因是：在线性空间中，用代数开集族来定义的拓扑与空间的线性结构不协调，即它不能使该空间因此成为拓扑线性空间. 但是如果我们用代数开凸集族来定义线性空间的拓扑基，那么这一拓扑就与空间的线性结构协调了. 在这一观点下，线性空间中的凸分析结果又变成几乎都能看成拓扑线性空间中的结果的特例. ▮

命题 1.3.2 如果 A 的代数内部 $A^i\neq\varnothing$，那么

$$\operatorname{aff} A=X.$$

证明 设 $a\in A^i$. 那么

$$\forall h\in X,\ \exists\varepsilon>0,\ [a,\ a+\varepsilon(h-a)]\subset A.$$

因此，$$b_h=a+\varepsilon(h-a)\in A.$$ 而

$$h=a+\frac{1}{\varepsilon}\cdot\varepsilon(h-a)=a+\frac{1}{\varepsilon}(b_h-a)\in\overline{ab_h}\subset\text{aff}\,A,$$

故 $h\in\text{aff}\,A$. 由 h 的任意性，即得

$$\text{aff}\,A=X. \blacksquare$$

命题1.3.3 n 维单纯形有非空相对代数内部.

证明 设 n 维单纯形

$$S=\Big\{x\in X\,|\,x=\lambda_0a_0+\lambda_1a_1+\cdots+\lambda_na_n,$$
$$\lambda_i\geqslant0,\ i=0,\ 1,\ \cdots,\ n,\ \sum_{i=0}^n\lambda_i=1\Big\},$$

这里 $a_0,\ a_1,\ \cdots,\ a_n\in X$ 为 $n+1$ 个仿射无关向量；特别是

$$e_1=a_1-a_0,\ \cdots,\ e_n=a_n-a_0$$

是 n 个线性无关向量. 在此指出，有下列形式的点

$$y=\tilde{\lambda}_0a_0+\tilde{\lambda}_1a_1+\cdots+\tilde{\lambda}_na_n,$$
$$\tilde{\lambda}_i>0\ \ (i=0,\ 1,\ \cdots,\ n),\ \sum_{i=0}^n\tilde{\lambda}_i=1,$$

是 S^{ri} 的元素. 事实上，令

$$\delta=\min_{0\leqslant i\leqslant n}\tilde{\lambda}_i>0.$$

则集合
$$G_y=\Big\{z\in X\,|\,z=y+\mu_1e_1+\cdots+\mu_ne_n,$$
$$|\mu_i|\leqslant\frac{\delta}{n},\ i=1,\ \cdots,\ n\Big\}\subset S.$$

这是因为
$$\begin{aligned}z&=y+\mu_1e_1+\cdots+\mu_ne_n\\&=(\tilde{\lambda}_0a_0+\tilde{\lambda}_1a_1+\cdots+\tilde{\lambda}_na_n)+\mu_1(a_1-a_0)\\&\quad+\mu_1(a_1-a_0)+\cdots+\mu_n(a_n-a_0)\\&=(\tilde{\lambda}_0-\mu_1-\cdots-\mu_n)a_0+(\tilde{\lambda}_1+\mu_1)a_1+\cdots\\&\quad+(\tilde{\lambda}_n+\mu_n)a_n;\end{aligned}$$

由 μ_i 的定义，立即可得 $z\in S$.

另一方面，对于任何 $h\in\text{lin}\{e_1,\ \cdots,\ e_n\}$，总存在 $\varepsilon>0$，使得

$$th\in G_y-y,\ \forall t\in[0,\ \varepsilon].$$

而易证
$$\operatorname{lin}\{e_1,\ \cdots,\ e_n\}=\operatorname{aff}S-y,$$
由此即得 $y\in S^{ri}$. ▌

综合命题 1.3.2 和 1.3.3，有以下定理.

定理 1.3.4 设 A 为 n 维空间 $\mathbf{R}^n$ 中的凸集. 那么
$$\operatorname{aff}A=\mathbf{R}^n$$
当且仅当
$$A^i\neq\varnothing.$$
特别是，任何有限维非空凸集的相对内部是非空的.

证明 只需证明必要性. 事实上，由 $\operatorname{aff}A=\mathbf{R}^n$ 可知 A 中包含 $n+1$ 个仿射无关的向量. 从而由于 A 是凸集，A 中必包含某个 n 维单纯形. 由命题 1.3.3, $A^i\neq\varnothing$. ▌

下列无限维空间的特征说明定理 1.3.4 中的有限维条件是本质的.

定理 1.3.5 线性空间 X 是无限维的充要条件为存在凸集 $A\subset X$，满足
$$\operatorname{aff}A=X,\ A^i=\varnothing.$$

证明 设 $\{e_i\}_{i\in I}$ 为 X 的 Hamel 基，这里指标集 I 为无限集. 那么任何 $x\in X$ 有表示式 $x=\sum_{i\in I}x^ie_i$, 这里 x^i 中只有有限个不为零. 令
$$A=\{x\in X\,|\,x=\sum_{i\in I}x^ie_i,\ x^i\geqslant0,\ \forall i\in I\}.$$
则 A 显然是 X 的凸集，且任何 $x\in X$ 都可表示为 A 中两个元素之差，从而由 $0\in A$, $\operatorname{aff}A=\operatorname{lin}A=X$. 但 $A^i=\varnothing$. 事实上，对于任意的 $x_0=\sum_{i\in I}x_0^ie_i\in A$, 总存在某个 $x_0^j=0$. 于是对于 $e_j\in X$, 对任何 $\varepsilon>0$, 都有 $x_0-\varepsilon e_j\notin A$. ▌

下面刻划凸集的代数内部.

定理 1.3.6 设 $A\subset X$ 为凸集. $A^{ri}\neq\varnothing$, $x_0\in A^{ri}$. 那么对于任何 $x_1\in A$,
$$[x_0,\ x_1)\subset A^{ri}, \tag{1.3.4}$$
且

$$A^{ri}=\bigcup_{x_1\in A\setminus\{x_0\}}[x_0,\ x_1)=\bigcup_{x_1\in A^{ri}\setminus\{x_0\}}[x_0,\ x_1);\qquad(1.3.5)$$

特别是 A^{ri} 是凸集.

证明 设 $x\in(x_0,\ x_1)$，则

$$x=x_0+\lambda(x_1-x_0),\ \lambda\in(0,\ 1).$$

由 $x_0\in A^{ri}$，故对任何 $h\in\operatorname{aff}A-A$，存在 $\varepsilon>0$，使得

$$[x_0,\ x_0+\varepsilon h]\subset A.$$

但由于 A 是凸集，故又有(如图 1.3)

$$x+(1-\lambda)\varepsilon h=(1-\lambda)(x_0+\varepsilon h)+\lambda x_1\in A,$$

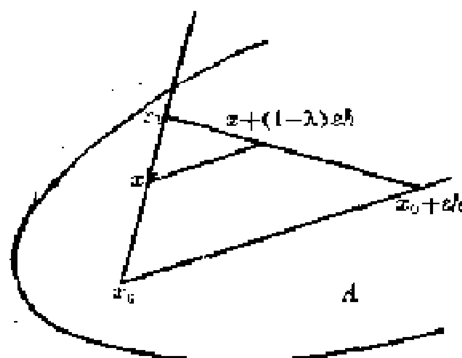


图 1.3

从而 $[x,\ x+(1-\lambda)\varepsilon h]\subset A$. 这样也就证得了 $x\in A^{ri}$. 由 x 的任意性，即得(1.3.4).

为证明(1.3.5)，只需指出

$$A^{ri}\subset\bigcup_{x_1\in A^{ri}\setminus\{x_0\}}[x_0,\ x_1).$$

事实上，设 $x\in A^{ri}$，那么对于 $h=x-x_0$，存在 $\delta>0$，使得

$$[x,\ x+2\delta(x-x_0)]\subset A;$$

从而由(1.3.4)，$x+\delta(x-x_0)\in A^{ri}$. 因此，

$$\begin{aligned}x=x_0+(x-x_0)\in\{y\in X\mid y&=x_0+\lambda(1+\delta)(x-x_0),\ \lambda\in[0,\ 1]\}\\&=[x_0,\ x+\delta(x-x_0))\subset\bigcup_{x_1\in A^{ri}\setminus\{x_0\}}[x_0,\ x_1),\end{aligned}$$

即(1.3.5)得证. ▮

推论 1 如果 $A\subset X$ 是凸集，那么 A^i(可能是 $\varnothing$)是凸的代数开集.

事实上，由(1.3.5)，当 $A^i\neq\varnothing$ 时，必定有 $(A^i)^i=A^i$. ▮

推论 2 如果凸集 A、B 满足 $A^{ri}\cap B^{ri}\neq\varnothing$，那么

$$A^{ri}\cap B^{ri}=(A\cap B)^{ri}.\qquad(1.3.6)$$

证明 事实上，设

$$x_0\in A^{ri}\cap B^{ri}\subset(A\cap B)^{ri},$$

则由(1.3.5)有

$$(A\cap B)^{ri}=\bigcup_{x_1\in(A\cap B)\setminus\{x_0\}}[x_0,\ x_1)=\{\bigcup_{x_1\in A\setminus\{x_0\}}[x_0,\ x_1)\}$$

$$\cap\{\bigcup_{x_1\in B\setminus\{x_0\}}[x_0,\ x_1)\}=A^{ri}\cap B^{ri}.\ ▮$$

注 条件 $A^{ri}\cap B^{ri}\neq\varnothing$ 是不可少的. 否则, 我们可取 A 为定理 1.3.5 中的 A, $B=-A$, 那么

$$A^{ri}=A^i=\varnothing,\quad B^{ri}=B^i=\varnothing,$$

但
$$(A\cap B)^{ri}=\{0\}^{ri}=\{0\}\neq\varnothing.\ ▮$$

下面引入代数闭包、代数闭集和代数边界等概念. 设 A 为线性空间X中的集合. 我们称

$$A^C=\{x\in X\mid \exists h\in X,\ \forall \varepsilon>0,\ \exists t\in[0,\ \varepsilon],\ x+th\in A\}\tag{1.3.7}$$

为 A 的代数闭包. 如果 $A^C=A$, 则称 A 为代数闭集. $A^b=A^c\setminus A$, 则称为 A 的代数边界.

比较 A^i 的定义(1.3.3)与 A^c 的定义(1.3.7), 可得下面的定理.

定理 1.3.7 对于任何集合 $A\subset X$,

$$X\setminus A^c=(X\setminus A)^i;\quad X\setminus A^i=(X\setminus A)^c.$$

特别是 A 为代数闭集当且仅当 $X\setminus A$ 为代数开集. ▮

注 1 在许多文献中, A^c 定义为

$$A^c=\{x\in X\mid \exists x_0\in A,\ [x_0,\ x)\subset A\}.$$

这一定义有许多不方便之处. 首先是定理 1.3.7 不成立; 其次是按照这一定义, $A^c\supset A$ 不一定成立, 于是连单点集也不是代数闭集. 我们的定义取自 C. Ursescu, SIAM J. Control and Optimization 20(1982), p. 563~574. 对于多于一点的凸集来说, 两种定义是一致的. ▮

注 2 如同代数开集一样, 现在的代数闭集也是有点奇怪的. 例如 $\mathbf{R}^2$ 中的一个圆周的任何子集都是代数闭集. ▮

命题 1.3.8 设 A 为线性空间 X 中的凸集. 那么 A^c 也是凸集.

证明 不妨设 A 多于一个点. 为指出 A^c 是凸集, 只需注意

到对于凸集 A，当 x, $y\in A^c$ 时，存在 x_0、$y_0\in A$，使得

$$[x_0, x), [y_0, y)\subset A.$$

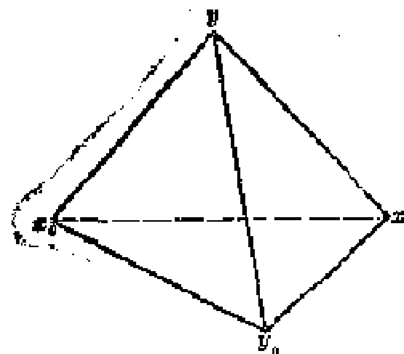


图 1.4

由于 A 是凸集，在单纯形 x_0y_0yx(如果它们在同一平面，情况也一样)中，可能除了棱 $[x, y]$、$(x_0, y]$、$[x, y_0)$ 以外，都必定是 A 中的元素(如图 1.4)．从而 $[x, y]$ 中的元素也必定是 A^c 的元素．▌

下列定理是定理 1.3.6 的加强．

定理 1.3.9　设 $A\subset X$ 为凸集，$A^{ri}\neq\varnothing$, $x_0\in A^{ri}$．那么对于任何 $x_1\in A^c$，有

$$[x_0, x_1)\subset A^{ri}. \tag{1.3.8}$$

特别是

$$A^c=\{x\in X\mid [x_0, x)\subset A^{ri}\}, \tag{1.3.9}$$

$$A^{ri}=\bigcup_{x_1\in A^c\setminus(x_0)}[x_0, x_1), \tag{1.3.10}$$

且 A^c 是凸的代数闭集．

证明　不妨设 $x_0=0$，且 A 多于一点．因为 $x_1\in A^c$，由 A 是凸集，故存在 $y\in A$，使得

$$[y, x_1)\subset A.$$

但由于 $0\in A^{ri}$，故有某个 $\varepsilon>0$，使得

$$[0, -\varepsilon y]\subset A.$$

于是对于任何 $t\in(0, 1)$，可求得

$$\delta\in(0, \varepsilon),$$

使得 $\overline{(-\delta y)(tx_1)}$ 与 $[y, x_1)$ 相交(如图 1.5)．这说明 $tx_1\in A$，从而由定理 1.3.6 可得

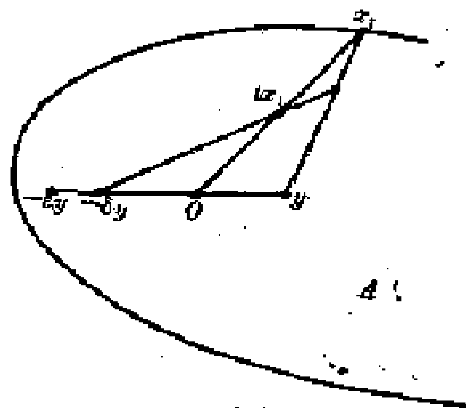


图 1.5

$$[0, tx_1)\subset A^{ri}, \ \forall t\in(0, 1),$$

以至 $[0, x_1)\subset A^{ri}$．即(1.3.8)对 $x_0=0$ 成立．

(1.3.9)由 A^c 的定义(1.3.7)和(1.3.8)可得．(1.3.10)也由(1.3.8)可得．最后，A^c 凸由命题 1.3.8 或(1.3.9)可得．而由(1.3.10)、(1.3.5)可得

$$(A^{c})^{ri} = A^{ri},$$

再由(1.3.9)可得

$$(A^{c})^{c} = A^{c}. \blacksquare$$

推论 对于任何有限维凸集 A, A^{c} 是代数闭集. ■

这是因为这时总有 $A^{ri} \neq \varnothing$.

下列定理说明条件 $A^{ri} \neq \varnothing$ 是必不可少的.

命题 1.3.10 线性空间 X 是无限维的充要条件为: 存在凸集 $A \subset X$, 使得

$$(A^{c})^{c} \neq A^{c}.$$

证明 设 X 为无限维线性空间. $\{e_i\}_{i \in I}$ 为其 Hamel 基. 令 $n(x)$ 为 $x = \sum_{i \in I} x^i e_i$ 中 $x^i \neq 0$ 的 i 的个数;

$$A = \{x \in X \mid x = \sum_{i \in I} x^i e_i,\ x^i \geqslant 0,\ \forall i \in I,\ \sum_{i \in I} x^i \geqslant 1/n(x)\}.$$

那么对于任何

$$x = \sum x^i e_i, \quad y = \sum y^i e_i \in A, \quad \lambda \in (0, 1),$$

有
$$n((1-\lambda)x + \lambda y) \geqslant \max(n(x), n(y)),$$

以致
$$\sum_{i \in I} \{(1-\lambda)x^i + \lambda y^i\} \geqslant \frac{1-\lambda}{n(x)} + \frac{\lambda}{n(y)} \geqslant \frac{1}{n((1-\lambda)x + \lambda y)}.$$

因此, A 是凸集.

另一方面, 可以指出:

$$A^{c} = \{x \in X \mid x = \sum_{i \in I} x^i e_i,\ x_i \geqslant 0,\ \forall i \in I,\ n(x) > 0\}. \tag{1.3.11}$$

事实上, 设 $z = \sum_{i \in I} z^i e_i$ 满足(1.3.11)右端, 则存在正整数 m, 使得

$$\sum_{i \in I} z^i \geqslant 1/m.$$

任取
$$e_{\alpha_1}, \cdots, e_{\alpha_m} \in \{e_i\}_{i \in I} \subset A,$$

则有
$$w = (e_{\alpha_1} + \cdots + e_{\alpha_m})/m \in A,$$

而再由
$$n((1-\lambda)w + \lambda z) \geqslant m, \quad \forall \lambda \in [0, 1),$$

立即可得 $[w, z) \subset A$. 因此, $z \in A^{c}$. 反之, 如果 $z \in A^{c}$, 则显然有 $n(z) \geqslant 0$; 但 $n(z) \neq 0$, 否则, 将存在

$$w=\sum_{i\in I}w^i e_i\in A,$$

使 $[w, 0)\subset A$. 然而这是不可能的，因为当

$$0<\lambda<\frac{1}{n(w)\sum w^i}$$

时，$\lambda w\notin A$. 这样，(1.3.11) 成立.

最后，显然有 $0\in(A^c)^c$. 因此

$$(A^c)^c\neq A^c. \quad ▌$$

作为这节的结束，我们再指出一个无限维空间的特征.

定理 1.3.11　线性空间 X 是无限维的充要条件为存在 X 的凸的真子集 A，使得

$$A^c=X.$$

证明　设 $X=\mathbf{R}^n$ 为有限维空间，凸集 A 满足 $A^c=X=\mathbf{R}^n$. 那么由 $\mathbf{R}^n=A^c\subset\text{aff}\,A$ 和定理 1.3.4, $A^i\neq\varnothing$. 再由定理 1.3.9 的(1.3.10)，有

$$A^i=\bigcup_{x_1\in A^c\setminus\{x_0\}}[x_0, x_1)=\bigcup_{x_1\in X\setminus\{x_0\}}[x_0, x_1)=X,$$

这里 $x_0\in A^i$. 因此，$A=X$，即 A 不是 X 的真子集.

反之，设 X 为无限维空间，$\{e_i\}_{i\in I}$ 为其 Hamel 基，这里 I 为无限集. 取可数个 $e_1, e_2, \cdots, e_n, \cdots\in\{e_i\}_{i\in I}$，令 $e_1, e_2, \cdots, e_n, \cdots$ 张成的线性子空间为 X_0，基中其余的元素张成的线性子空间为 X_1(可能是零维的). 那么

$$X=X_0+X_1.$$

令 $A_0\subset X_0$ 为所有其最后一个坐标大于零的向量全体，即 A_0 的元素

$$x_0=\sum_{n=1}^{\infty}x_0^n e_n$$

的有限个不为零的 x_0^n 中，n 最大的那个 $x_0^n>0$. 取

$$A=A_0+X_1,$$

则 A 显然是凸集，且 $A\neq X$.

但另一方面，任取

$$x=\sum_{n=1}^{m}x^n e_n+x_1\in X,$$

这里 $x_1\in X_1$，如果它满足当 $n\geqslant m$ 时，$x^n=0$，则由于 $e_{m+1}\in X_0$，且对于任何 $t>0$，有

$$x+te_{m+1}\in A,$$

故 $x\in A^c$. 由于 x 是任取的，即得

$$A^c=X. \blacksquare$$

注 线性空间 X 的子集 A 若满足 $A^c=X$，则称 A 为 X 的弥漫集(ubiquitous). 定理 1.3.11 说明，当且仅当 X 无限维时，X 才有凸的弥漫真子集. 弥漫集显然也就是 X 用代数开集族赋以拓扑时的稠密集. 不过这是由于对 A^c 采用了前面提到的 Ursesco 的定义的缘故：如果采用常用的定义，弥漫集就不具有拓扑意义.

还要注意的是：凸的弥漫真子集的代数内部一定是空的. 这由定理 1.3.9 的(1.3.10)就可看出. $\blacksquare$

§1.4 凸集分离定理

前面曾经提到过一类重要的仿射集——超平面：

$$H=\{x\in X\mid\langle x^*, x\rangle=\alpha\},\quad x^*\in X^*,\quad x^*\neq 0,\quad \alpha\in\mathbf{R}.$$

一个超平面把空间 X 分成两个部分，即

$$H_-^0=\{x\in X\mid\langle x^*, x\rangle<\alpha\};$$

$$H_+^0=\{x\in X\mid\langle x^*, x\rangle>\alpha\}.$$

它们分别称为由超平面 H 确定的两个开半空间. 容易验证，这两个开半空间都是代数开凸集. 类似地，我们称

$$H_-^c=\{x\in X\mid\langle x^*, x\rangle\leqslant\alpha\};$$

$$H_+^c=\{x\in X\mid\langle x^*, x\rangle\geqslant\alpha\}$$

为由超平面 H 确定的两个闭半空间；它们都是代数闭凸集. 上述四个半空间构成两组互补的凸集，即

$$H_+^c\cup H_-^0=X,\quad H_+^c\cap H_-^0=\varnothing;$$

$$H_+^0\cup H_-^c=X,\quad H_+^0\cap H_-^c=\varnothing.$$

一般来说，当两个凸集 C、D 满足

$$C\cup D=X,\quad C\cap D=\varnothing$$

时，就称它们为互补的凸集.

X 中的两个子集 A、B 称为被超平面 H 所分离，是指它们分别落在由 H 所确定的两个闭半空间中，例如，$A\subset H^c_-$，$B\subset H^c_+$；或者说，存在 $x_0^*\in X^*$，$x_0^*\neq 0$，满足

$$\langle x_0^*, x\rangle\leqslant\langle x_0^*, y\rangle,\ \forall x\in A,\ \forall y\in B. \tag{1.4.1}$$

如果(1.4.1)换为

$$\langle x_0^*, x\rangle<\langle x_0^*, y\rangle,\ \forall x\in A,\ \forall y\in B, \tag{1.4.2}$$

那么 A、B 称为被超平面 H 严格分离[*)]. 如果再进一步要求

$$\sup_{x\in A}\langle x_0^*, x\rangle<\inf_{y\in B}\langle x_0^*, y\rangle, \tag{1.4.3}$$

那么 A、B 称为被超平面 H 强分离. 可以看出，当 A、B 被超平面 H 分离时，它们在 H 上还可能有公共点；当 A、B 被严格分离时，它们不能再有公共点. 但 A^c、B^c 还可能有公共点；而当 A、B 被强分离时，连 A^c、B^c 都不再有公共点.

从平面上的图形可以看出，两个不相交的平面凸集总能用直线(一维超平面)把它们分离. 这个结论实际上对任何线性空间上的任何具有非空相对代数内部的凸集都是成立的，尤其是在任何有限维空间中总是成立的. 这类结果称为凸集分离定理；它们是凸分析理论的主要工具. 其证明方法很多，在此利用超锥的概念来证明它.

首先指出，任何两个集合的分离等价于它们的代数差与原点的分离，即

命题 1.4.1 线性空间 X 中的集合 A、B 可用超平面分离(严格分离、强分离)的充要条件为：0 可以与集合 $A-B$ 用超平面分离(严格分离，强分离).

证明 事实上，(1.4.1)等价于

---

*) 注意，这个定义不等价于

$$\langle x_0^*, x\rangle<\alpha<\langle x_0^*, y\rangle,\ \forall x\in A,\ \forall y\in B,$$

而有些文献却是这样来定义严格分离的.

$$\langle x_0^*, x-y\rangle \leqslant 0, \ \forall x\in A, \ \forall y\in B,$$

而上式正说明了 $A-B$ 与 0 可用超平面分离. 严格分离、强分离情形的证明是一样的. ■

由命题1.4.1, 我们只需讨论一个集合与 0 的分离问题. 我们的思路大致如下: 设凸集K不包含 0, 作K的凸锥包,然后把这凸锥扩大到不能再大(但小于全空间)为止. 最后的这个极大凸锥称为超锥. 当K的相对代数内部非空时, 这个超锥的边界就是使K与 0 分离的超平面. 这里就需要超锥的概念, 而为了得到包含K的超锥, 则需要 Zorn 引理, 最后为使超锥的边界是超平面, 需要一个相对代数内部非空的条件.

线性空间X中的不包含顶点 0 的凸锥 C 称为超锥, 是指不存在包含 C 作为真子集的不包含顶点 0 的凸锥.

命题 1.4.2 凸锥 C 为超锥的充要条件为

$$C\cup(-C)=X\setminus\{0\}, \quad C\cap(-C)=\varnothing.$$

证明 设凸锥 C 为超锥,则必须有

$$C\cap(-C)=\varnothing.$$

否则,由 C 是凸的可得 $0\in C$. 同时,又必须有

$$C\cup(-C)=X\setminus\{0\}.$$

否则,存在 $x_0\neq 0$, $x_0\notin C\cup(-C)$, 以致射线 $\overrightarrow{0x_0}\notin C\cup(-C)$. 于是 $0\notin\overrightarrow{0x_0}+C$, 且 $\overrightarrow{0x_0}+C$ 是比 C 更大的凸锥. 这与 C 是超锥相矛盾.

反之,如果

$$C\cup(-C)=X\setminus\{0\}, \ C\cap(-C)=\varnothing,$$

则不存在 $x_0\neq 0$, 使得

$$0\notin\overrightarrow{0x_0}+C,$$

从而不存在不包含 0 的比 C 更大的凸锥. ■

推论 1 C 为超锥的充要条件是: $-C$ 也为超锥. ■

推论 2 如果 C 为超锥,那么令

$$D=(-C)\cup\{0\},$$

则 C 与 D 为互补的凸集. ■

命题 1.4.3 设凸集K不包含0. 那么存在包含K的超锥.

证明 设$\mathscr{C}$为包含K而不包含0的凸锥的全体. 那么由于

$$\bigcup_{\lambda>0}\lambda K\in\mathscr{C},$$

故$\mathscr{C}$非空. 同时在$\mathscr{C}$中以包含关系为序关系时, 满足 Zorn 引理的条件; 因此, 由 Zorn 引理, $\mathscr{C}$中有极大元, 而该极大元显然必须是超锥. ▌

命题 1.4.4 设C与D为X中的两个非空互补凸集,

$$M=C^c\cap D^c.$$

那么或者M是X中的超平面, 或者$M=X$.

证明 因为C与D是凸集, 由命题 1.3.8, C^c与D^c也是凸集, 以致$M=C^c\cap D^c$也是凸集. 我们再进一步指出M是仿射集. 事实上, 由C与D互补, 可得

$$C^c=X\setminus D^i,\quad D^c=X\setminus C^i$$

(定理 1.3.7). 因此,

$$M=(X\setminus D^i)\cap(X\setminus C^i)=X\setminus(C^i\cup D^i).$$

若存在x、$y\in M$, $z\in\overline{xy}$, 但$z\notin M$, 则$z\in C^i\cup D^i$. 不妨设$z\in C^i$, 且$y\in[z,x)$. 那么由定理 1.3.9, $y\in C^i$, 这与$y\in M$矛盾. 这样M是仿射集.

现在不妨假设M是线性子空间. 如果$M\neq X$, 那么存在

$$p\in X\setminus M=C^i\cup D^i.$$

由于$0\in M$, 则必须有$-p\in X\setminus M$; 否则, 将有

$$p\in\overline{0(-p)}\subset M.$$

不妨设$p\in C^i$, 则必须有$-p\in D^i$; 否则, 将有$-p\in C^i$, 以致

$$0\in[-p,\ p]\in C^i,$$

这与$0\in M$相矛盾.

这样一来, 我们有

$$p\in C^i,\ -p\in D^i,\ M=C^c\cap D^c.$$

从而由定理 1.3.9, 易证

$$[-p,\ x]\cap M\neq\varnothing,\ \forall x\in C;$$

$$[p, y]\cap M\neq\varnothing, \quad \forall y\in D.$$

(如图 1.6). 特别是，这说明

$$X=C\cup D=\text{lin}(\{p\}\cup M).$$

最后，由命题 1.2.1，可把 M 上的零线性形式延拓到全空间 X，并使它在 p 上为 1；即存在 $x^*\in X^*$，满足

$$\langle x^*, p\rangle=1, \ \langle x^*, x\rangle=0, \quad \forall x\in M.$$

因此，M 是超平面. ∎

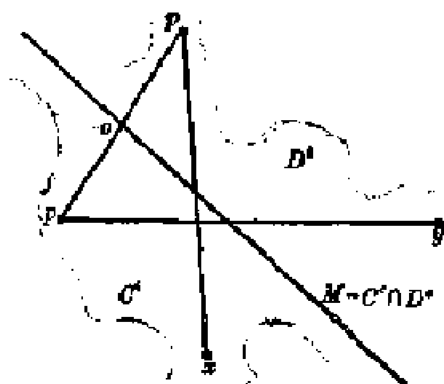


图 1.6

注 $M=X$ 是可能的. 例如，取 C 为定理 1.3.11 证明中的弥漫集 A_0(易证它是 X_0 的超锥)，$D=(-A_0)\cup\{0\}$. 则

$$M=A_0^c\cap\{(-A_0)^c\cup\{0\}\}=X.$$

下面证明本节的主要结果：

定理 1.4.5 设线性空间 X 中的凸集 K 的代数内部 $K^i\neq\varnothing$，且 $0\notin K^i$. 那么可用超平面将 0 与 K 分离，与 K^i 严格分离.

证明 只需证明，可用超平面将 0 与 K^i 严格分离. 事实上，由 K^i 是凸集，$0\notin K^i$，故根据命题 1.4.3，存在包含 K^i 的超锥 C. 令

$$D=(-C)\cup\{0\}.$$

则 C 与 D 为两个非空互补凸集，且由命题 1.4.4，$M=C^c\cap D^c$ 只可能是超平面或全空间 X. 如果 M 是超平面，由 $K^i\subset C^i$ 可知，M 严格分离 0 与 K^i. 我们再指出 M 不可能是全空间 X，否则，由

$$M=C^c\cap D^c=X$$

可知 $C^c=X$，即 C 是弥漫集. 但 $K^i\subset C^i$，$K^i\neq\varnothing$，C 又是 X 的凸真子集. 因此，这是不可能的. ∎

定理 1.4.5 可改进为下面的定理：

定理 1.4.5′ 设凸集 K 的相对代数内部 $K^{ri}\neq\varnothing$，且 $0\notin K^{ri}$. 那么可用超平面将 0 与 K 分离，与 K^{ri} 严格分离.

证明 如果 $0\notin \text{aff}\,K$，那么 $\text{lin}(K-K)$ 为 X 的真子空间，它是 $\text{aff}\,K$ 的平移，即

$$\operatorname{aff} K = x_0 + \operatorname{lin}(K-K),$$

其中 $x_0 \notin \operatorname{lin}(K-K)$. 由命题 1.2.1, 存在 $x_0^* \in X^*$, 满足

$$\langle x_0^*, x_0\rangle = 1; \quad \langle x_0^*, x\rangle = 0, \ \forall x \in \operatorname{lin}(K-K).$$

这样, 超平面

$$H = \{x \in X \mid \langle x_0^*, x\rangle = 0\}$$

严格分离(甚至强分离)0 与 K.

如果 $0 \in \operatorname{aff} K$, 那么

$$X_1 = \operatorname{aff} K = \operatorname{lin} K$$

为X的子空间, 且K在 X_1 上的代数内部非空. 在子空间 X_1 上利用定理 1.4.5, 则可得到 $x_1^* \in X_1^*$ 满足

$$\langle x_1^*, x\rangle < 0, \ \forall x \in K^{ri}.$$

由命题 1.2.1, x_1^* 可延拓到全空间 X, 且使上式仍成立. ▌

利用命题 1.4.1 立即可得以下定理.

定理 1.4.6(凸集分离定理) 设 A、B 为线性空间X中的两个集合. 如果 $A-B$ 为凸集,

$$(A-B)^{ri} \neq \varnothing,$$

且 $0 \notin (A-B)^{ri}$, 那么 A、B 可用超平面分离, 0 与$(A-B)^{ri}$ 可用超平面严格分离. ▌

注 出于习惯,我们仍称定理 1.4.6 为凸集分离定理,但在定理 1.4.6 中实际上只要求 $A-B$ 为凸集. 下面的例子说明, 当 $A-B$ 为有非空代数内部的凸集时, A、B 都可以不是凸集, 也都可以没有代数内部.

例 令 $X = \mathbf{R}^2$.

$$A = \{(x^1, x^2) \mid (x^1-1)^2 + (x^2)^2 = 1\};$$

$$B = \{(x^1, x^2) \mid (x^1+1)^2 + (x^2)^2 = 1\}.$$

那么 $A-B = A+A = \{(x^1, x^2) \mid (x^1-2)^2 + (x^2)^2 \leqslant 2^2\}$

(如图 1.7).

推论 1 设 A、B 为两个不相交的有限维非空凸集. 那么 A、B 可用超平面分离. ▌

因为有限维凸集的相对内部总是非空的.

推论 2 设 A、B 为两个非空凸集，$A^i\neq\varnothing$，且
$A^i\cap B=\varnothing$。

那么 A、B 可用超平面分离，A^i、B 可用超平面严格分离.

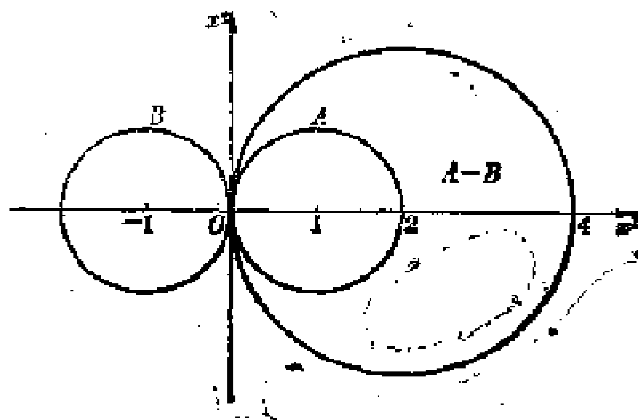


图 1.7

证明 事实上，这时 A^i 为非空代数开凸集，从而对于任何 $x\in X$，A^i-x 也是非空代数开凸集. 因此，由 B 凸以及命题 1.3.1,

$$A^i-B=\bigcup_{x\in B}(A^i-x)$$

也是非空代数开凸集. 这样,

$$A^i-B=(A^i-B)^i\neq\varnothing,$$

且由 $A^i\cap B=\varnothing$，而有 $0\notin(A^i-B)$. ▌

推论 3(凸集承托定理) 设 A 为 X 的凸集，且 $A^{ri}\neq\varnothing$. 那么对于任何 $x\notin A^{ri}$，存在 $x^*\in X^*$，使得

$$\langle x^*, x\rangle<\langle x^*, y\rangle,\ \forall y\in A^{ri}. ▌$$

只需考虑凸集 $A-x$，并注意到

$$(A-x)^{ri}=A^{ri}-x.$$

所谓凸集 A 被超平面 H 所承托，是指 A 落在 H 所确定的一个闭半空间中，且

$$H\cap A\neq\varnothing.$$

如果还要求 $A\not\subset H$，那么称 H 真承托 A，$H\cap A$ 中的元素称为承托点. 推论 3 说明 A 的每个"相对代数边界点"($x\in A\backslash A^{ri}$)，都是 A 的真承托点.

推论 2 不能再改进为 $A^{ri}\neq\varnothing$. 否则，由于单点集的相对代数内部是非空的，会使定理 1.4.5′ 没有条件 $K^{ri}\neq\varnothing$，也可成立. 但是利用下列命题可使推论 2 得到一定程度的推广.

命题 1.4.7 设 A、B 为线性空间 X 中的任意集，且 $A^{ri}\neq\varnothing$,

$B^{ri}\neq\varnothing$. 那么

$$A^{ri}+B^{ri}\subset(A+B)^{ri};\qquad(1.4.4)$$

如果 A、B 还是凸集，那么

$$A^{ri}+B^{ri}=(A+B)^{ri}.\qquad(1.4.5)$$

证明 设 $x_1\in A^{ri}$, $x_2\in B^{ri}$. 我们证明

$$x_1+x_2\in(A+B)^{ri}.$$

事实上，任取

$$h\in\operatorname{aff}(A+B)-(x_1+x_2),$$

那么 h 可表示为

$$h=\lambda_1(y_1+z_1)+\cdots+\lambda_n(y_n+z_n)-x_1-x_2,$$

这里 $\lambda_1+\cdots+\lambda_n=1$, $y_1,\ \cdots,\ y_n\in A$, $z_1,\ \cdots,\ z_n\in B$.

令
$$h_1=\sum_{i=1}^n\lambda_iy_i-x_1\in\operatorname{aff}A-x_1;$$

$$h_2=\sum_{i=1}^n\lambda_iz_i-x_2\in\operatorname{aff}A-x_2.$$

则由 $x_1\in A^{ri}$, $x_2\in B^{ri}$, 存在 $\varepsilon>0$, 使得

$$[x_1,\ x_1+\varepsilon h_1]\subset A,\quad[x_2,\ x_2+\varepsilon h_2]\subset B.$$

因此，
$$[x_1+x_2,\ x_1+x_2+\varepsilon(h_1+h_2)]$$
$$\subset[x_1,\ x_1+\varepsilon h_1]+[x_2,\ x_2+\varepsilon h_2]\subset A+B.$$

由 $h=h_1+h_2$ 的任意性，$x_1+x_2\in(A+B)^{ri}$, 从而(1.4.4)得证.

如果 A、B 是凸集，则 $A+B$ 也是凸集. 根据定理 1.3.6, $(A+B)^{ri}$ 由形为 $[x_1+x_2,\ y_1+y_2)$ 的区间构成. 这里

$$x_1\in A^{ri},\ y_1\in A,\ x_2\in B^{ri},\ y_2\in B.$$

但
$$[x_1+x_2,\ y_1+y_2)\subset[x_1,\ y_1)+[x_2,\ y_2)\subset A^{ri}+B^{ri},$$

这样(1.4.5)成立. ∎

定理1.4.8 设 A、B 为线性空间 X 中的凸集，

$$A^{ri}\neq\varnothing,\quad B^{ri}\neq\varnothing,$$

且
$$A^{ri}\cap B^{ri}=\varnothing.$$

那么，A、B 可用超平面分离；A^{ri}、B^{ri} 可用超平面严格分离. ∎

事实上，这时会有

$$0\notin A^{ri}-B^{ri}=(A-B)^{ri}.$$

于是定理 1.4.8 归结为定理 1.4.6.

还可注意，定理 1.4.6 也可看作 A 退化为点 0，B 代替为 $A-B$ 时的定理 1.4.8 的特殊情形.

现在讨论强分离的条件. 基本结果如下：

定理 1.4.9 设 A、B 为 X 中的两个非空集合，$A-B$ 为凸集，$(A-B)^{ri}\neq\varnothing$，且 $0\notin(A-B)^c$. 那么，A、B 可用超平面强分离.

证明 如果 $0\notin\mathrm{aff}(A-B)$，则已在定理 1.4.5′ 的证明中指出了 0 与 $A-B$ 可用超平面强分离.

现在设 $0\in\mathrm{aff}(A-B)$. 取 $x_0\in(A-B)^{ri}$，连结 0 和 x_0，则由 $0\notin(A-B)^c$，存在 $x_1\in(0,\ x_0)$，使得

$$[x_1,\ x_0)\subset(A-B)^c,\quad [0,\ x_1)\subset X\setminus(A-B)^c.$$

这样也有 $x_1\notin(A-B)^{ri}$，故由定理 1.4.6 的推论 3，存在 $x_1^*\in X^*$，满足

$$\langle x_1^*,\ x\rangle<\langle x_1^*,\ x_1\rangle,\ \forall x\in(A-B)^{ri},\tag{1.4.6}$$

以致

$$\sup_{x\in(A-B)^c}\langle x_1^*,\ x\rangle\leqslant\langle x_1^*,\ x_1\rangle.\tag{1.4.7}$$

但由 $x_1\in(0,\ x_0)$，可知 $x_1=\lambda_0x_0$，这里 $\lambda_0\in(0,\ 1)$. 因此，由 $x_0\in(A-B)^{ri}$ 和 (1.4.6) 可得

$$\langle x_1^*,\ x_0\rangle<\langle x_1^*,\ x_1\rangle=\lambda_0\langle x_1^*,\ x_0\rangle,$$

从而 $\langle x_1^*,\ x_0\rangle<0$，以致 $\langle x_1^*,\ x_1\rangle<0$. 这样，由 (1.4.7) 有

$$\sup_{x\in(A-B)^c}\langle x_1^*,\ x\rangle<0,$$

即 0 与 $(A-B)^c$ 可用超平面强分离，特别是 A、B 可用超平面强分离. ▌

注 我们不能指望定理 1.4.9 也有类似定理 1.4.8 那样的推广，即认为当

$$A^{ri}\neq\varnothing,\ B^{ri}\neq\varnothing,\ A^c\cap B^c=\varnothing$$

时，A、B 即可用超平面强分离. 这是因为不可能有

$$A^c+B^c=(A+B)^c$$

这样的一般等式. 下面是一个反例:

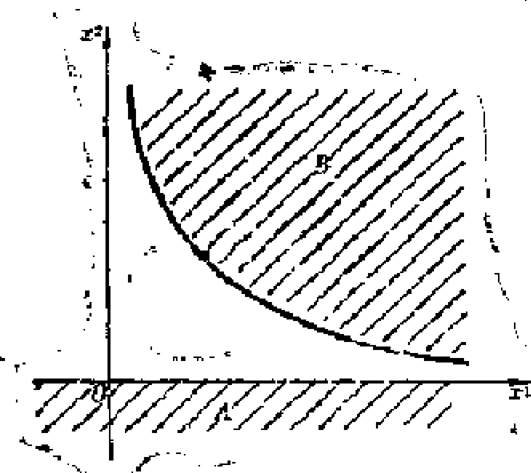


图　1.8

$X=\mathbf{R}^2$, $A=\{(x^1, x^2)\mid x^2\leqslant 0\}$,

$$B=\left\{(x^1, x^2)\mid x^2\geqslant \frac{1}{x^1}, x^1>0\right\};$$

但 A、B 不能用超平面强分离(如图 1.8). ▌

下面是另一个强分离判定定理. 为此先提出吸收集的概念. X 中的集合 V 称为吸收集, 是指 $0\in V^i$. 它也可直接定义为 V 满足条件

$$\forall x\in X, \exists\lambda>0, \forall\alpha\geqslant\lambda, x\in\alpha V.$$

定理 1.4.10　X 中的两个非空凸集 A、B 可用超平面强分离的充要条件为: 存在 X 中的吸收凸集 V, 使得

$$(A+V)\cap B=\varnothing.$$

证明　如果这样的 V 存在, 那么 $A+V$ 是凸集, 且由

$$A\subset(A+V)^i$$

可得

$$(A+V)^i\neq\varnothing.$$

根据定理 1.4.6 的推论 2, 存在 $x^*\in X^*$, $x^*\neq 0$, 使得

$$\sup_{x\in B}\langle x^*, x\rangle\leqslant\inf_{y\in A, z\in V}\langle x^*, y+z\rangle. \tag{1.4.8}$$

但由 $x^*\neq 0$, 故存在 $x_0\in X$, 使得

$$\langle x^*, x_0\rangle<0.$$

由 V 是吸收集, 存在 $\alpha>0$, 使得

$$x_1=x_0/\alpha\in V,$$

从而也有 $\langle x^*, x_1\rangle<0$. 这样, 由 (1.4.8) 可得

$$\sup_{x\in B}\langle x^*, x\rangle\leqslant\inf_{y\in A}\langle x^*, y\rangle+\langle x^*, x_1\rangle<\inf_{y\in A}\langle x^*, y\rangle,$$

即 A、B 被强分离.

反之, 如果存在 $x^*\in X^*$, 使得

$$\sup_{x\in B}\langle x^*, x\rangle<\inf_{y\in A}\langle x^*, y\rangle,$$

那么令
$$\varepsilon=\frac{1}{2}\{\inf_{y\in A}\langle x^*, y\rangle-\sup_{x\in B}\langle x^*, x\rangle\},$$
$$V=\{x\in X\,|\,|\langle x^*, x\rangle|<\varepsilon\},$$
易证 V 是吸收凸集，且
$$(A+V)\cap B=\varnothing. \quad ▌$$

推论 如果 A、B 为两个非空凸集，
$$(A-B)^{ri}\neq\varnothing,$$
且
$$0\notin(A-B)^{o},$$
那么存在吸收凸集，使得
$$(A+V)\cap B=\varnothing. \quad ▌$$

第一章习题

1. 设 H 是线性空间 X 的超平面，$A\supset H$ 是 X 的仿射集。试证明：或者 $A=H$，或者 $A=X$。
2. 设 X、Y 为两个线性空间，A: $X\to Y$ 为线性映射，$A\subset X$，$B\subset Y$ 为两个空间中的凸集。试证明：
$$A(A)=\{y\in Y\,|\,\exists x\in A,\ Ax=y\};$$
$$A^{-1}(B)=\{x\in X\,|\,\exists y\in B,\ Ax=y\}$$
都是凸集。
3. 设 A 为 n 维空间中的集合，那么 $\text{lin}\,A$($\text{cone}\,A$)中的每个点都至多是 A 中的 n 个点的线性组合(锥组合)；$\text{aff}\,A$($\text{co}\,A$)中的每个点都至多是 A 中的 $n+1$ 个点的仿射组合(凸组合)(Carathéodory 定理)。
4. 试证明：如果 A 是代数开集，那么 A 的凸包 $\text{co}\,A$ 也是代数开集。
5. 试证明：如果 A 是锥，那么 A^i 也是锥。
6. 设 A: $X\to Y$ 为线性空间 X 到线性空间 Y 的线性映射，且 $A(X)=Y$。如果 $A\subset X$ 是代数开集，试证明：$A(A)\subset Y$ 也是代数开集。
7. 试证明：如果 A、B 为 X 中的两个不相交的凸集，那么存在两个互补凸集 C、D，使得 $A\subset C$，$B\subset D$（角谷静夫-Stone 定理)。
8. 试利用上题的结果证明凸集分离定理。
9. 设 A 为非空仿射集，B 为凸集，且 $B^{ri}\neq\varnothing$，$A\cap B^{ri}=\varnothing$。试证明：存在包含 A 的超平面，使 B 在该超平面的一侧(即在一个闭半空间中)。
10. 设 A 为凸集，$A^{ri}\neq\varnothing$，B 为非空凸多面集(有限个闭半空间之交，可以

不封闭)，$A\cap B=\varnothing$．试证明：存在超平面H严格分离 A、B.

11．设 C 为线性空间 X 中的非空代数开凸锥．试证明：存在 $x_0^*\in X^*$，使得

$$\langle x_0^*, x\rangle<0,\ \ \forall x\in C.$$

12．设 C_1、C_2 为线性空间 X 中的两个凸锥，且

$$C_1^{ri}\neq\varnothing,\ C_2^{ri}\neq\varnothing,\ C_1^{ri}\cap C_2^{ri}=\varnothing.$$

试证明：存在非零 x_1^*、$x_2^*\in X^*$，使得

$$\langle x_1^*, x\rangle\geqslant 0\ \ \forall x\in C_1;$$

$$\langle x_2^*, x\rangle\geqslant 0\ \ \forall x\in C_2.$$

而 $x_1^*+x_2^*=0$．试把这个结果推广到有限个凸锥的情形．

13．试在 $\mathbf{R}^2$ 中构造凸集 A、B，使得

$$A^c+B^c\neq(A+B)^c.$$

14．设 A、B 为线性空间 X 中的非空集合，且满足

$$(1-\lambda)A+\lambda B\subset A,\ \forall\lambda\in[0, 1).$$

试证明：$A^c=B^c$，$A^i=B^i$.

15．在定理1.4.9．的证中 $A^{ri}\neq\varnothing$，$B^{ri}\neq\varnothing$，$A^c\cap B^c=\varnothing$ 并不足以保证凸集 A、B 可用超平面强分离．试指出这样的条件是否足以保证凸集 A、B 可用超平面严格分离？分析以下的例子：

$$X=\mathbf{R}^3,\ A=\{(x, y, z)\mid x\geqslant 0,\ y\geqslant 0,\ z\geqslant 0,\ z^2\leqslant xy\},$$

$$B=\{(x, y, z)\mid x=0,\ z=1\}.$$

第二章　凸函数与凸规划

§2.1　单变量凸函数

在讨论一般线性空间上的凸函数以前，先在这节中讨论实数区间上的单变量凸函数.

设$(a, b)\subset \boldsymbol{R}$为实数轴上的开区间，$-\infty\leqslant a<b\leqslant+\infty$. 函数$f$: $(a, b)\to\boldsymbol{R}$称为(a, b)上的凸函数,是指

$$\left.\begin{array}{l} f((1-\lambda)x_1+\lambda x_2)\leqslant(1-\lambda)f(x_1)+\lambda f(x_2), \\ \forall x_1、x_2\in(a, b), \ \forall\lambda\in[0, 1]. \end{array}\right\} \quad (2.1.1)$$

它的几何意义为其图象是下凸的(如图 2.1).

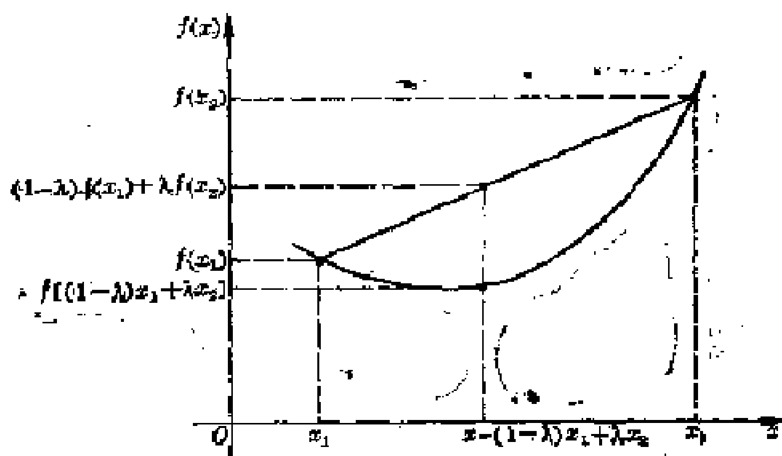


图　2.1

如果(2.1.1)中的不等号始终是严格的，那么f称为(a, b)上的严格凸函数. 如果$-g$是(a, b)上的凸函数,那么g称为(a, b)上的凹函数. 此外,还可注意,(2.1.1)也等价于更强的不等式

$$f\left(\sum_{i=1}^{n}\lambda_i x_i\right)\leqslant\sum_{i=1}^{n}\lambda_i f(x_i),$$

$$\forall x_1, \cdots, x_n\in(a, b), \ \lambda_1, \cdots, \lambda_n\in[0, 1], \ \sum_{i=1}^{n}\lambda_i=1.$$

命题 2.1.1　f为(a, b)上的凸函数等价于下列条件中的任

何一个:

i) $\forall x_1、x_2\in(a,\ b),\ x_2>x_1,\ \forall x\in(x_1,\ x_2)$,

$$\frac{f(x_1)-f(x)}{x_1-x}\leqslant\frac{f(x_2)-f(x)}{x_2-x} \tag{2.1.2}$$

即左差商不大于右差商.

ii) $\forall x_1、x_2\in(a,\ b),\ x_2>x_1,\ \forall x\in(x_1,\ x_2)$,

$$\frac{f(x)-f(x_1)}{x-x_1}\leqslant\frac{f(x_2)-f(x_1)}{x_2-x_1} \tag{2.1.3}$$

即右差商当自变量差分减小时是不增的.

iii) $\forall x_1、x_2\in(a,\ b),\ x_2>x_1,\ \forall x\in(x_1,\ x_2)$,

$$\frac{f(x_1)-f(x_2)}{x_1-x_2}\leqslant\frac{f(x)-f(x_2)}{x-x_2} \tag{2.1.4}$$

即左差商当自变量差分减小时是不减的.

iv) $F(x,\ y)=\dfrac{f(x)-f(y)}{x-y},\ x\neq y$, 作为 x 或 y 的函数在 $(a,\ b)$ 上不减.

证明 设 $x=(1-\lambda)x_1+\lambda x_2$. 则(2.1.2)等价于

$$\frac{f(x_1)-f(x)}{\lambda(x_1-x_2)}\leqslant\frac{f(x_2)-f(x)}{(1-\lambda)(x_2-x_1)}$$

即 $\qquad f(x)=f((1-\lambda)x_1+\lambda x_2)\leqslant(1-\lambda)f(x_1)+\lambda f(x_2)$.

(2.1.3)、(2.1.4)的证明类似, iv)只是 i)~iii)的统一叙述. ∎

命题 2.1.2 设 f 为 $(a,\ b)$ 上的凸函数. 那么 f 在 $(a,\ b)$ 上处处左、右可导, 从而处处连续. 同时, 其左、右导数 f'_-、f'_+ 满足:

$$\forall x_1,\ x_2\in(a,\ b),\ x_1<x_2,$$

$$f'_-(x_1)\leqslant f'_+(x_1)\leqslant\frac{f(x_2)-f(x_1)}{x_2-x_1}$$

$$\leqslant f'_-(x_2)\leqslant f'_+(x_2). \tag{2.1.5}$$

证明 事实上, 由(2.1.2)与(2.1.3)可知, 当 $x>x_2>x_1$ 时, 有

$$f'_+(x_2)=\lim_{\substack{x\to x_2\\ x>x_2}}\frac{f(x)-f(x_2)}{x-x_2}$$

$$=\inf_{x>x_2}\frac{f(x)-f(x_2)}{x-x_2}\geqslant\frac{f(x_2)-f(x_1)}{x_2-x_1}. \tag{2.1.6}$$

同样，由(2.1.2)与(2.1.4)可知，当 $x<x_1<x_2$ 时，有

$$f'_-(x_1)=\lim_{\substack{x\to x_1\\x<x_1}}\frac{f(x)-f(x_1)}{x-x_1}$$

$$=\sup_{x<x_1}\frac{f(x)-f(x_1)}{x-x_1}\leqslant\frac{f(x_2)-f(x_1)}{x_2-x_1}. \qquad (2.1.7)$$

由 x_1、x_2 的任意性，(2.1.6)和(2.1.7)同时也指出了 f 在 (a, b) 中处处左、右可导．再由(2.1.2)－(2.1.4)易证，当 $x_1<x_2$ 时，有下式成立：

$$f'_+(x_2)\geqslant f'_-(x_2)\geqslant\frac{f(x_2)-f(x_1)}{x_2-x_1};$$

$$\frac{f(x_2)-f(x_1)}{x_2-x_1}\geqslant f'_+(x_1)\geqslant f'_-(x_1). \quad \blacksquare$$

命题 2.1.2 的逆也成立．其证明需要下列推广的中值定理：

引理 2.1.3 设 (a, b) 上的连续函数 f 处处有右导数 f'_+．那么

$$\inf_{x\in(x_1,x_2)}f'_+(x)\leqslant\frac{f(x_2)-f(x_1)}{x_2-x_1}\leqslant\sup_{x\in(x_1,x_2)}f'_+(x),$$

$$\forall x_1,\ x_2\in(a,\ b).$$

证明 首先证明，如果 (x_1, x_2) 上的连续函数 $g(x)$ 处处有 g'_+，且

$$g'_+(x)\geqslant 0,\quad \forall x\in(x_1,\ x_2), \qquad (2.1.8)$$

那么

$$\forall x'_1、x'_2\in(x_1,\ x_2),\ x'_2>x'_1,\ g(x'_2)\geqslant g(x'_1). \qquad (2.1.9)$$

事实上，这时有

$$\forall x\in(x_1,\ x_2),\ \forall\varepsilon>0,\ \exists\delta_x>0,\ \forall h\in[0,\ \delta_x],$$
$$g(x+h)-g(x)\geqslant-\varepsilon h.$$

于是设

$$x'=\sup\{x\in[x'_1,\ x'_2]\mid g(x)-g(x'_1)\geqslant-\varepsilon(x-x'_1)\}.$$

那么必须有 $x'=x'_2$．否则，有 $x'<x'_2$，从而由 x' 的定义和 $g(x)$ 的连续性，可知

$$g(x')-g(x'_1)\geqslant-\varepsilon(x'-x'_1).$$

而对于 x'，又存在 $\delta_{x'}>0$，使得

$$\forall h\in[0,\ \delta_{x'}],\ g(x'+h)-g(x')\geqslant -\varepsilon h,$$

以至
$$g(x'+h)-g(x_1')\geqslant -\varepsilon(x'+h-x_1'),$$

这与 x' 的定义相矛盾．这样一来，就有

$$g(x_2')-g(x_1')\geqslant -\varepsilon(x_2'-x_1').$$

由 ε 的任意性，(2.1.9)得证．

现在令
$$M=\sup_{x\in(x_1,x_2)}f'_+(x).$$

那么
$$g(x)=Mx-f(x)$$

满足(2.1.8)．因此，由(2.1.9)，当

$$x_1',\ x_2'\in(x_1,\ x_2),\ x_2'>x_1'$$

时，
$$Mx_2'-f(x_2')\geqslant Mx_1'-f(x_1').$$

由 f 的连续性，即得

$$\frac{f(x_2)-f(x_1)}{x_2-x_1}\leqslant M=\sup_{x\in(x_1,x_2)}f'_+(x).$$

再令
$$m=\inf_{x\in(x_1,x_2)}f'_+(x),$$

对
$$h(x)=f(x)-mx$$

作同样讨论，引理即得证．▌

定理 2.1.4　f 是$(a,\ b)$上的凸函数的充要条件为 f 在$(a,\ b)$处处左、右可导，且其左、右导数 f'_-、f'_+ 满足

$$\forall x_1,\ x_2\in(a,\ b),\ x_1<x_2,$$

$$f'_-(x_1)\leqslant f'_+(x_1)\leqslant\frac{f(x_2)-f(x_1)}{x_2-x_1}$$

$$\leqslant f'_-(x_2)\leqslant f'_+(x_2).\qquad(2.1.5)$$

证明　只需证明充分性．事实上，由引理 2.1.3 和(2.1.5)，立即可得

$$\forall x_1,\ x_2\in(a,\ b),\ x_2>x_1,\ \forall x\in(x_1,\ x_2),$$

$$\frac{f(x_1)-f(x)}{x_1-x}\leqslant\sup_{y\in(x_1,x)}f'_+(y)\leqslant\inf_{y\in(x,x_2)}f'_+(y)\leqslant\frac{f(x_2)-f(x)}{x_2-x}.$$

因此，由命题 2.1.1，f 为$(a,\ b)$上的凸函数．▌

推论 1　设 f 为 $(a,\ b)$ 上的可导函数．那么 f 是$(a,\ b)$上的

凸函数的充要条件为 $f'(x)$ 在 (a, b) 上不减. ▌

推论 2 设 f 为 (a, b) 上的二次可导函数. 那么 f 是 (a, b) 上的凸函数的充要条件为

$$f''(x) \geqslant 0, \ \forall x \in (a, b). \text{▌}$$

推论 3 设 f 为 (a, b) 上的凸函数. 那么 f 在 (a, b) 上至多除可数个点外, 处处可导.

证明 f 的不可导点即

$$f'_-(x_0) \neq f'_+(x_0)$$

的点 $x_0 \in (a, b)$. 由(2.1.5)可知, 在 f 的任何两个不可导点 x_0, $x_1 \in (a, b)$, 区间 $(f'_-(x_0), f'_+(x_0))$ 和 $(f'_-(x_1), f'_+(x_1))$ 不相交. 从而这种区间至多只有可数个. ▌

注 1 把上面的论证中的不等号改为严格不等号, 即可相应地得到一系列有关严格凸函数的结果. ▌

注 2 用定义直接来判断一个函数是不是凸函数, 往往是很困难的. 但是用定理 2.1.4 的推论 1 和 2 来判断一个光滑函数是否凸, 则是相当简便的. 在实际应用中, 常常先用导数来肯定函数的凸性. 然而, 反过来引出它必定满足凸性不等式. 例如, 利用二阶导数, 可以肯定 $y = x^\alpha$ 当 $\alpha \geqslant 1$ 时是 $(0, \infty)$ 上的凸函数. 因此, 下列不等式成立:

$$(\lambda_1 x_1 + \cdots + \lambda_n x_n)^\alpha \leqslant \lambda_1 x_1^\alpha + \cdots + \lambda_n x_n^\alpha,$$

$$\forall x_1, \cdots, x_n \geqslant 0, \ \forall \lambda_1, \cdots, \lambda_n \in [0, 1], \ \lambda_1 + \cdots + \lambda_n = 1;$$

特别是 $\left(\dfrac{x_1 + \cdots + x_n}{n}\right)^\alpha \leqslant \dfrac{x_1^\alpha + \cdots + x_n^\alpha}{n}, \ \forall x_1, \cdots, x_n \geqslant 0.$

如所周知, 这类不等式的直接证明是不容易的. ▌

注 3 尽管凸函数总是至多除了可数个点以外, 处处可导, 但它的导数却可能在一个处处稠密的集合上不存在; 例如, 设 $\{r_n\}$ 为 $\boldsymbol{R}$ 上的有理数全体,

$$\alpha_n > 0 \ (n = 1, 2, \cdots), \ \sum_{n=1}^{\infty} \alpha_n < \infty.$$

令

$$g(x) = \sum_{r_n < x} \alpha_n, \ \forall x \in \boldsymbol{R}.$$

$$f(x)=\int_0^x g(t)dt, \ \forall x\in \boldsymbol{R}.$$

那么，由于 g 为递增函数，f 为凸函数．但 f 在所有有理点上都不存在导数．由于这样的函数存在，我们一般无法讨论凸函数的二阶导数．然而，А. Д. Александров(Уч. записки ЛГУ. Сер. матем., 6(1936), 3-35)指出，凸函数却几乎处处存在二阶逼近，即

$$f(x)=f(x_0)+\frac{f'(x_0)}{1!}(x-x_0)+\frac{\tilde{f}''(x_0)}{2!}(x-x_0)^2$$
$$+o(|x-x_0|^2),$$

于是也可以说，凸函数几乎处处有上式意义下的"二阶导数" $\tilde{f}''(x_0)$．∎

在这一节的最后，我们再把函数的凸性与图象空间中的凸集联系起来．

设 $f:(a, b)\to \boldsymbol{R}$ 为区间 (a, b) 上的函数．我们称图象空间 $\boldsymbol{R}^2$ 中的集合

$$\operatorname{epi} f=\{(x, y)\in \boldsymbol{R}^2 \mid x\in(a, b), f(x)\leqslant y\}$$

为 f 的上图(epigraph)．

命题 2.1.5 f 在 (a, b) 上凸等价于 $\operatorname{epi} f$ 是 $\boldsymbol{R}^2$ 中的凸集．

证明 设 f 为 (a, b) 上的凸函数，(x_1, y_1)、$(x_2, y_2)\in \operatorname{epi} f$．则

$$x_1, x_2\in(a, b), f(x_1)\leqslant y_1, f(x_2)\leqslant y_2.$$

由(2.1.1)，

$$f((1-\lambda)x_1+\lambda x_2)\leqslant(1-\lambda)f(x_1)+\lambda f(x_2)$$
$$\leqslant(1-\lambda)y_1+\lambda y_2, \ \forall\lambda\in[0, 1],$$

即 $$(1-\lambda)(x_1, y_1)+\lambda(x_2, y_2)\in \operatorname{epi} f, \ \forall\lambda\in[0, 1].$$

因此，$\operatorname{epi} f$ 为凸集．

反之，设 $\operatorname{epi} f$ 为凸集．则由于对于任何

$$x_1、x_2\in(a, b), (x_1, f(x_1)), (x_2, f(x_2))\in \operatorname{epi} f,$$

故 $$(1-\lambda)(x_1, f(x_1))+\lambda(x_2, f(x_2))\in \operatorname{epi} f, \ \forall\lambda\in[0, 1].$$

因此，

$$f((1-\lambda)x_1+\lambda x_2)\leqslant(1-\lambda)f(x_1)+\lambda f(x_2), \ \forall\lambda\in[0, 1],$$

即 f 是凸函数. ∎

命题 2.1.6 设 f 为 (a, b) 上的凸函数. 那么对于任何 $x_0 \in (a, b)$, 存在 $\alpha \in \boldsymbol{R}$, 使得

$$f(x)-f(x_0) \geqslant \alpha(x-x_0), \ \forall x \in (a, b).$$

证明 由命题 2.1.5, $\operatorname{epi} f$ 是凸集, 而

$$(x_0, f(x_0)) \notin (\operatorname{epi} f)^i \neq \varnothing.$$

因此, 由凸集分离定理 1.4.8, 存在 $(\alpha_1, \alpha_2) \in \boldsymbol{R}^{2*}$, 使得

$$\alpha_1 x+\alpha_2 y<\alpha_1 x_0+\alpha_2 f(x_0), \ \forall (x, y) \in (\operatorname{epi} f)^i.$$

因为 y 可任意大, 故为使上述不等式成立, 必须有 $\alpha_2 \leqslant 0$. 但 α_2 也不能为零, 否则, 由

$$(x_0, f(x_0)+3) \in (\operatorname{epi} f)^i,$$

上述不等式不可能成立. 因此, $\alpha_2<0$. 取 $\alpha=\alpha_1/-\alpha_2$, 即得

$$y-f(x_0)>\alpha(x-x_0), \ \forall (x, y) \in (\operatorname{epi} f)^i.$$

由于 y 可任意靠近 $f(x)$, 故命题得证. ∎

由命题的证明可见, 这里的 α 就是凸集 $\operatorname{epi} f$ 在点 $(x_0, f(x_0))$ 的承托直线的斜率.

引入下列记号:

$$\partial f(x_0)=\{\alpha \in \boldsymbol{R} \mid f(x)-f(x_0) \geqslant \alpha(x-x_0), \ \forall x \in (a, b)\},$$

并称它为函数 f 在点 x_0 的次微分. 命题 2.1.6 说明, 对于凸函数来说, $\partial f(x_0)$ 总是非空的.

命题 2.1.7 设 f 为 (a, b) 上的凸函数. 那么

$$\partial f(x_0)=[f'_-(x_0), f'_+(x_0)], \ \forall x_0 \in (a, b).$$

证明 设

$$\alpha \in [f'_-(x_0), f'_+(x_0)].$$

则当 $x>x_0$, 由 (2.1.5) 可得

$$\frac{f(x)-f(x_0)}{x-x_0} \geqslant f'_+(x_0) \geqslant \alpha;$$

当 $x<x_0$, 由 (2.1.5) 可得

$$\frac{f(x)-f(x_0)}{x-x_0} \leqslant f'_-(x_0) \leqslant \alpha,$$

即
$$f(x)-f(x_0)\geqslant\alpha(x-x_0),\ \forall x\in(a,\ b).$$

反之，如果 $\alpha\in\partial f(x_0)$，则同样由(2.1.5)，有

$$f'_+(x_0)=\inf_{x>x_0}\frac{f(x)-f(x_0)}{x-x_0}\geqslant\alpha\geqslant\sup_{x<x_0}\frac{f(x)-f(x_0)}{x-x_0}$$
$$=f'_-(x_0),$$

即
$$\alpha\in[f'_-(x_0),\ f'_+(x_0)].\ ▮$$

命题2.1.7说明，次微分 ∂f 这个概念可看作导数概念的推广．凸函数 f 在点 x_0 可导当且仅当 $\partial f(x_0)$ 只包含一点．以后我们将把这一概念推广到一般情形．

作为这节的结束，我们还应指出，(2.1.1)也可作为闭区间 $[a,\ b]$ 上的凸函数的定义，即允许 x_1、x_2 取 a 或 b．闭区间上的凸函数在区间内部的性态当然仍如前面所述的一样．但是在区间的端点上，凸函数不但可以没有单边导数，甚至可以不连续．然而，仍可指出，它必须在端点上上半连续，即

$$f(a)\geqslant\limsup_{x\to a^+}f(x),\ f(b)\geqslant\limsup_{x\to b^-}f(x).$$

其证明留给读者作为练习．

§2.2　线性空间上的凸函数

现在讨论线性空间上的凸函数．由于线性空间中只有代数结构，我们不能在其中讨论凸函数的连续性和可微性，从而在上节中单变量凸函数的大部分性质都只能在引进拓扑后的线性空间中才能推广．这将在第三章中叙述．本节中所论及的只是凸函数的“代数性质”．

设 K 为线性空间 X 中的一个凸集．K 上的函数 $f:K\to\mathbf{R}$ 称为凸函数，是指对于任何

$$x_1、x_2\in K,\ g(t)=f(x_1+t(x_2-x_1))$$

是 $t\in[0,\ 1]$ 上的凸函数．它也可直接定义为

$$\forall x_1、x_2\in K,\ \forall\lambda\in[0,\ 1],$$
$$f((1-\lambda)x_1+\lambda x_2)\leqslant(1-\lambda)f(x_1)+\lambda f(x_2).\quad(2.2.1)$$

同样也可定义严格凸函数与凹函数. 与以前一样, (2.2.1) 也等价于更强的不等式:

$$\forall x_1, \cdots, x_n \in K, \ \forall \lambda_1, \cdots, \lambda_n \in [0, 1], \ \sum_{i=1}^{n} \lambda_i = 1,$$

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leqslant \sum_{i=1}^{n} \lambda_i f(x_i). \tag{2.2.2}$$

线性空间上的凸函数也同样联系着图象空间 $X \times \mathbf{R}$ 中的一个凸集. 设 $f: K \to \mathbf{R}$ 为 $K \subset X$ 上的函数. 那么 $X \times \mathbf{R}$ 的子集

$$\operatorname{epi} f = \{(x, \alpha) \in X \times \mathbf{R} ; x \in K, \ f(x) \leqslant \alpha\} \tag{2.2.3}$$

称为 f 的上图. 于是同样有

命题 2.2.1 f 是凸集 $K \subset X$ 上的凸函数当且仅当 f 的上图 $\operatorname{epi} f$ 为 $X \times \mathbf{R}$ 的凸集. ▮

证明与单变量的情形完全一样.

利用上图的概念和命题 2.2.1, 我们也可以用 "$\operatorname{epi} f$ 是凸集" 来作为凸函数的定义. 这一新定义的好处还在于: 它可把凸函数的概念推广到取广义实值 ($\mathbf{R} \cup \{\pm\infty\}$) 的函数. 对于取广义实值的凸函数用定义 (2.2.1), 会遇到 $(+\infty)+(-\infty)$ 等不定运算, 但利用上图则不会有此困难. 不过, 如果对 $\pm\infty$ 的运算, 除了作通常的规定:

$$\forall \alpha \in \mathbf{R}, \quad (+\infty)+\alpha = +\infty, \ (-\infty)+\alpha = -\infty,$$

$$(+\infty)+(+\infty) = +\infty, \ (-\infty)+(-\infty) = -\infty,$$

$$\forall \lambda > 0, \quad \lambda(+\infty) = +\infty, \ -\lambda(+\infty) = -\infty,$$

$$0 \cdot (+\infty) = 0 \cdot (-\infty) = 0,$$

此外, 再规定

$$(+\infty)+(-\infty) = +\infty, \tag{2.2.4}$$

那么, 不难验证, 用 (2.2.1) 来作为取广义实值的凸函数的定义, 仍有命题 2.2.1 成立.

把凸函数的值域扩充到 $\mathbf{R} \cup \{\pm\infty\}$ 以后, 还可看到任何凸集 K 上的凸函数 f 都可用下列方式延拓到全空间 X, 而成为 X 上的凸函数:

$$\tilde{f}(x)=\begin{cases} f(x), & \text{当 } x\in K; \\ +\infty, & \text{当 } x\notin K. \end{cases}$$

这样一来，以后只需讨论全空间 X 上的取广义实值的凸函数．具体地说，函数

$$f: X\to \boldsymbol{R}\cup\{\pm\infty\}$$

为 X 上的凸函数，是指其上图

$$\operatorname{epi} f=\{(x,\ \alpha)\in X\times\boldsymbol{R}\mid f(x)\leqslant\alpha\} \tag{2.2.5}$$

为 $X\times\boldsymbol{R}$ 中的凸集，或者

$$\forall x_1,\ x_2\in X,\ \forall\lambda\in[0,\ 1],$$

$$f((1-\lambda)x_1+\lambda x_2)\leqslant(1-\lambda)f(x_1)+\lambda f(x_2), \tag{2.2.6}$$

其中对 $\pm\infty$ 的加法运算，除遵循常用的规定的外，还遵循(2.2.4)．

注　需注意的是，对等式(2.2.4)只许用“消去”，即

$$(+\infty)+(-\infty)+\alpha=+\infty+\alpha,\quad \forall\alpha\in\boldsymbol{R}\cup\{\pm\infty\},$$

但不许用“移项”，即不准把(2.2.4)的两端中的项移到另一端去．例如，把右端移到左端得到

$$(+\infty)+(-\infty)-(+\infty)=0,$$

或把左端的一项移到右端得到

$$-\infty=+\infty-(+\infty)$$

等等．▌

对于凸函数 $f: X\to\boldsymbol{R}\cup\{\pm\infty\}$，下列 X 的集合

$$\operatorname{dom} f=\{x\in X\mid f(x)<\infty\}$$

称为 f 的**有效域**；不取 $-\infty$，且

$$\operatorname{dom} f\neq\varnothing$$

的凸函数称为**真凸函数**．真凸函数实际上是真正有意义的凸函数．如果 f 是凸函数，但不是真凸函数，那么或者 $f(x)\equiv+\infty$；或者存在 $x_0\in\operatorname{dom} f$，使得 $f(x_0)=-\infty$．而在后一情形，对于任何 $x\in(\operatorname{dom} f)^{ri}$，存在 $\varepsilon>0$，使得

$$x_1=x+\varepsilon(x-x_0)\in\operatorname{dom} f,$$

从而
$$x=\frac{1}{1+\varepsilon}x_1+\frac{\varepsilon}{1+\varepsilon}x_0.$$
因此，由(2.2.6)，
$$f(x)\leqslant\frac{1}{1+\varepsilon}f(x_1)+\frac{\varepsilon}{1+\varepsilon}f(x_0)=-\infty.$$
这样，f 只能在 $\mathrm{dom}\,f$ 的相对代数边界点
$$x\in\mathrm{dom}\,f\setminus(\mathrm{dom}\,f)^{ri}$$
上取有限值.

命题 2.2.2 设 f_i $(i=1,\cdots,n)$都是线性空间 X 上的凸函数. 那么，对于任何 $\lambda_i\geqslant0(i=1,\cdots,n)$，函数
$$f=\sum_{i=1}^n\lambda_if_i:\quad x\mapsto\sum_{i=1}^n x_if_i(x),\ \forall x\in X$$
也是X上的凸函数，其中加法运算遵循(2.2.4). ∎

命题 2.2.3 设 f_i $(i\in I)$都是线性空间X上的凸函数. 那么函数
$$f=\sup_{i\in I}f_i:\quad x\mapsto\sup_{i\in I}f_i(x),\ \forall x\in X \tag{2.2.7}$$
也是X上的凸函数. 特别是，仿射函数(线性形式与常数函数之和)族的上包络是凸函数，这里上包络是指用(2.2.7)方式来定义的函数. ∎

证明由定义即得. 命题 2.2.2 适于用(2.2.6)来证明，而命题 2.2.3 则适于用上图来证明，因为下列等式成立:
$$\mathrm{epi}(\sup_{i\in I}f_i)=\bigcap_{i\in I}\mathrm{epi}\,f_i.$$

在此指出，命题 2.2.3 的后半部分的逆在一定条件下也成立，即在一定条件下，凸函数可表示为仿射函数族的上包络.

命题 2.2.4 设 f 为线性空间X上的真凸函数，
$$(\mathrm{dom}\,f)^{ri}\neq\varnothing,$$
$A(f)$表示所有不大于 f 的仿射函数全体，即 $a\in A(f)$ 表示存在 $x_a^*\in X^*$，$\alpha_a\in\boldsymbol{R}$，使得
$$a(x)=\langle x_a^*,x\rangle+\alpha_a\leqslant f(x),\quad\forall x\in X.$$
那么

$$f(x)=\sup_{a\in A(f)} a(x),\ \forall x\in(\mathrm{dom}\ f)^{ri}. \tag{2.2.8}$$

证明　此命题的证明与命题2.1.6的证明类似．事实上，只需证明，对于任何 $x_0\in(\mathrm{dom}\ f)^{ri}$，存在 $a_{x_0}\in A(f)$，满足

$$a_{x_0}(x_0)=f(x_0).$$

首先，由 $(\mathrm{dom}\ f)^{ri}\neq\varnothing$，可指出 $(\mathrm{epi}\ f)^{ri}\neq\varnothing$．这是因为可以验证

$$(\mathrm{epi}\ f)^{ri}=\{(x,\ \alpha)\in X\times\boldsymbol{R}\,|\,x\in(\mathrm{dom}\ f)^{ri},\ f(x)<\alpha\}. \tag{2.2.9}$$

为此，设 $(x,\ \alpha)$ 满足 $x\in(\mathrm{dom}\ f)^{ri}$，$f(x)<\alpha$．那么对于任何

$$(y,\ \beta)\in\mathrm{aff\ epi}\ f-(x,\ \alpha)\subset(\mathrm{aff\ dom}\ f-x)\times\boldsymbol{R},$$

存在 $\delta>0$，使得

$$x+\delta y\in\mathrm{dom}\ f,$$

以至对于任何 $s\in(0,\ 1)$，有

$$\begin{aligned}&f(x+s\delta y)-(\alpha+s\delta\beta)\\&\quad=f((1-s)x+s(x+\delta y))-(\alpha+s\delta\beta)\\&\quad\leqslant(1-s)f(x)+sf(x+\delta y)-(1-s)\alpha-s(\alpha+\delta\beta)\\&\quad=(1-s)(f(x)-\alpha)+sc.\end{aligned}$$

这里 c 与 s 无关．上式右端当 s 充分小时是负的，从而有

$$\exists s>0,\ [(x,\ \alpha),\ (x,\ \alpha)+s\delta(y,\ \beta)]\subset\mathrm{epi}\ f,$$

即

$$(x,\ \alpha)\in(\mathrm{epi}\ f)^{ri}.$$

此外，$\mathrm{epi}\ f$ 中其他的元素显然不在 $(\mathrm{epi}\ f)^{ri}$ 中，因此，(2.2.9)成立．

其次，由 $(x_0,\ f(x_0))\notin(\mathrm{epi}\ f)^{ri}$，故由凸集分离定理1.4.8，存在 $(z^*,\ \gamma^*)\in X^*\times\boldsymbol{R}^*$，满足

$$\langle z^*,\ x_0\rangle+\gamma^*f(x_0)\leqslant\langle z^*,\ x\rangle+\gamma^*\alpha,\ \forall(x,\ \alpha)\in\mathrm{epi}\ f; \tag{2.2.10}$$

$$\langle z^*,\ x_0\rangle+\gamma^*f(x_0)<\langle z^*,\ x\rangle+\gamma^*\alpha,\ \forall(x,\ \alpha)\in(\mathrm{epi}\ f)^{ri}. \tag{2.2.11}$$

由这两个不等式，必须有 $\gamma^*\geqslant0$，否则，由 α 可任意大，上述两式不可能成立．同时，$\gamma^*=0$ 也不可能，否则，由

$$(x_0, f(x_0)+\varepsilon) \in (\text{epi} f)^{ri}, \quad \varepsilon>0,$$

也使(2.2.11)不成立. 这样, $\gamma^*>0$. 令

$$a_{x_0}(x) = -\langle z^*, x\rangle/\gamma^* + \langle z^*, x_0\rangle/\gamma^* + f(x_0). \quad (2.2.12)$$

则由(2.2.10), 将有

$$a_{x_0}(x) \leqslant f(x), \quad \forall x \in \text{dom} f.$$

因此, $a_{x_0} \in A(f)$, 且由(2.2.12),

$$a_{x_0}(x_0) = f(x_0). \blacksquare$$

推论 设 $f: X \to \boldsymbol{R}$ 为 X 上的凸函数. 那么

$$f(x) = \sup_{a \in A(f)} a(x), \quad \forall x \in X. \blacksquare$$

命题 2.2.4 及其推论说明了凸函数实质上结构是很简单的. 但是即使对于 $(\text{dom} f)^{ri} \neq \varnothing$ 的真凸函数, (2.2.8)也不能推广为

$$f(x) = \sup_{a \in A(f)} a(x), \quad \forall x \in \text{dom} f \text{ 或 } X.$$

例如, $X = \boldsymbol{R}$ 上的真凸函数:

$$f(x) = \begin{cases} 0, & \text{当 } x \in (-1, 1); \\ 1, & \text{当 } x = \pm 1; \\ +\infty, & \text{当 } |x| > 1. \end{cases}$$

那么在 $x=\pm 1$ 时, 上述等式是不可能成立的. 下列定理指出了等式成立的充要条件.

定理 2.2.5 设 $f: X \to \boldsymbol{R} \cup \{+\infty\}$ 为 X 上的任意函数,

$$\text{dom} f = \{x \in X \mid f(x) < +\infty\}, \quad (\text{dom} f)^{ri} \neq \varnothing.$$

那么, 存在一族仿射函数 a_i, $i \in I$, 使得

$$f(x) = \sup_{i \in I} a_i(x), \quad \forall x \in X \quad (2.2.13)$$

的充要条件为: epi f 是代数闭凸集.

证明 如果(2.2.13)成立, 则

$$\text{epi} f = \bigcap_{i \in I} \text{epi}\, a_i.$$

但每一 epi a_i 都是闭半空间, 所以 epi f 是代数闭凸集.

反之, 设 epi$f$ 是代数闭凸集. 由假定 $(\text{dom} f)^{ri} \neq \varnothing$, 可同上面一样证明

$$(\text{epi} f)^{ri} \neq \varnothing,$$

则对于任何 $x_0\in\operatorname{dom}f$ 和 $\varepsilon>0$，由

$$(x_0,\ f(x_0)-\varepsilon)\notin\operatorname{epi}f=(\operatorname{epi}f)^c,$$

利用凸集强分离定理1.4.9，可求得 $(z^*,\ \gamma^*)\in X^*\times\boldsymbol{R}^*$，使得

$$\langle z^*,\ x_0\rangle+\gamma^*(f(x_0)-\varepsilon)<\langle z^*,\ x\rangle+\gamma^*\alpha,\ \forall(x,\ \alpha)\in\operatorname{epi}f.$$

同前面一样讨论，仍可得到 $\gamma^*>0$．令

$$a_{x_0,\varepsilon}(x)=-\langle z^*,\ x\rangle/\gamma^*+\langle z^*,\ x_0\rangle/\gamma^*+(f(x_0)-\varepsilon),$$

于是 $a_{x_0,\varepsilon}\in A(f)$，且

$$a_{x_0,\varepsilon}(x_0)=f(x_0)-\varepsilon.$$

由 ε 的任意性，这就证明了

$$f(x)=\sup_{a\in A(f)}a(x),\quad\forall x\in\operatorname{dom}f.$$

还需指出，当 $f(x_1)=+\infty$ 时，也有

$$\sup_{a\in A(f)}a(x_1)=+\infty.$$

事实上，这时 $x_1\notin\operatorname{dom}f$，从而对于任何 $\beta\in\boldsymbol{R}$，$(x_1,\ \beta)\notin\operatorname{epi}f$．再利用定理1.4.9，可求得 $(z_1^*,\ \gamma_1^*)\in X^*\times\boldsymbol{R}^*$，使得

$$\langle z_1^*,\ x_1\rangle+\gamma_1^*\beta<\langle z_1^*,\ x\rangle+\gamma_1^*\alpha,\ \forall(x,\ \alpha)\in\operatorname{epi}f.$$

与前面一样，我们仍有 $\gamma_1^*\geqslant0$．如果始终有 $\gamma_1^*>0$，那么由于 β 可任意大，就有

$$\sup_{a\in A(f)}a(x_1)=+\infty.$$

但是现在可能有 $\gamma_1^*=0$．这时，利用强分离条件，则有

$$\inf_{x\in\operatorname{dom}f}\langle z_1^*,\ x-x_1\rangle=\delta>0.$$

或

$$0\geqslant\delta-\langle z_1^*,\ x-x_1\rangle,\ \forall x\in\operatorname{dom}f.\tag{2.2.14}$$

由上面的证明可知，$A(f)\neq\varnothing$，故存在

$$a(x)=\langle x_0^*,\ x\rangle+\gamma_0^*\in A(f),$$

且

$$f(x)\geqslant\langle x_0^*,\ x\rangle+\gamma_0^*,\ \forall x\in X.\tag{2.2.15}$$

把(2.2.15)的两端加上 n 倍的(2.2.14)的两端，于是有

$$f(x)\geqslant\langle x_0^*,\ x\rangle+\gamma_0^*+n\delta-\langle nz_1^*,\ x-x_1\rangle,\ \forall x\in\operatorname{dom}f.$$

令
$$a_{x_1,n}(x)=\langle x_0^*,\ x\rangle+\gamma_0^*+n\delta-\langle nz_1^*,\ x-x_1\rangle.$$

则 $a_{x_1,n}\in A(f)$，且

$$a_{x_1,n}(x_1)=\langle x_0^*,\ x_1\rangle+\gamma_0^*+n\delta.$$

因为 n 可任意大，这就证明了

$$\sup_{a\in A(f)} a(x_1)=+\infty. \blacksquare$$

在这节的最后证明下面与线性形式延拓有关的著名定理：

定理 2.2.6(Hahn–Banach) 设 L 为线性空间 X 的子空间，f 是 X 上的真凸函数，且

$$(\mathrm{dom}\ f)^{ri}\cap L\neq\varnothing.$$

如果 L 上的线性形式 $x_L^*\in L^*$，满足

$$\forall x\in L,\ \langle x_L^*,\ x\rangle\leqslant f(x), \tag{2.2.16}$$

那么存在 $x^*\in X^*$，满足

$$\forall x\in L,\ \langle x_L^*,\ x\rangle=\langle x^*,\ x\rangle; \tag{2.2.17}$$

$$\forall x\in X,\ \langle x^*,\ x\rangle\leqslant f(x). \tag{2.2.18}$$

证明 设

$$A=\mathrm{epi}\ f=\{(x,\ \alpha)\in X\times\boldsymbol{R}\,|\,f(x)\leqslant\alpha\};$$

$$B=\{(x,\ \alpha)\in X\times\boldsymbol{R}\,|\,x\in L,\ \langle x_L^*,\ x\rangle=\alpha\}. \tag{2.2.19}$$

那么 A 是凸集，且正如前面已指出的

$$A^{ri}=\{(x,\ \alpha)\,|\,x\in(\mathrm{dom}\ f)^{ri},\ f(x)<\alpha\}\neq\varnothing; \tag{2.2.20}$$

B 是 $X\times\boldsymbol{R}$ 上的子空间，从而

$$B^{ri}=B\neq\varnothing.$$

再由(2.2.16)和(2.2.19)、(2.2.20)可得

$$A^{ri}\cap B^{ri}=\varnothing.$$

这样一来，由定理 1.4.8，A^{ri}、$B^{ri}=B$ 可用超平面严格分离，即存在

$$(z^*,\ \gamma^*)\in X^*\times\boldsymbol{R}^*,$$

使得

$$\left.\begin{array}{l}\langle z^*,\ y\rangle+\gamma^*\langle x_L^*,\ y\rangle<\langle z^*,\ x\rangle+\gamma^*\alpha,\\ \forall y\in L,\ \forall(x,\ \alpha)\in A^{ri}.\end{array}\right\} \tag{2.2.21}$$

由 L 是子空间，故如果 $y\neq 0$，$y\in L$，则也有 $\forall\lambda\in\boldsymbol{R}$，$\lambda y\in L$，以至(2.2.21)左端必须恒为零，即

$$\langle z^*, y\rangle+\gamma^*\langle x_L^*, y\rangle=0, \qquad \forall y\in L, \tag{2.2.22}$$

$$\langle z^*, x\rangle+\gamma^*\alpha>0, \qquad \forall (x, \alpha)\in A^{ri}. \tag{2.2.23}$$

另一方面，由于(2.2.23)中 α 可任意大，故必须有 $\gamma^*\geqslant 0$.

而又由于 $(\mathrm{dom}\, f)^{ri}\cap L\neq\varnothing$，不可能有 $\gamma^*=0$，否则，(2.2.22)和(2.2.23)不能同时成立．因此，$\gamma^*>0$．最后，令

$$x^*=-z^*/\gamma^*.$$

由(2.2.22)和(2.2.23)容易验证(2.2.17)和(2.2.18)成立．▮

§2.3　次线性函数和 Minkowski 函数

有一类特别重要的凸函数 $f: X\to \boldsymbol{R}\cup\{+\infty\}$，称为*次线性函数*．它满足：

$$\left.\begin{array}{ll} \text{i)}\ \forall x\in X,\ \forall t\geqslant 0,\ f(tx)=tf(x); & (\text{正齐次性}) \\ \text{ii)}\ \forall x_1, x_2\in X,\ f(x_1+x_2)\leqslant f(x_1)+f(x_2). & (\text{次可加性}) \end{array}\right\} \tag{2.3.1}$$

任何线性形式(函数)当然都是次线性函数．反之，易证：如果 $f: X\to\boldsymbol{R}$，且 f 和 $-f$ 都是次线性函数，那么 f 一定是线性函数．

由定义(2.2.6)式出发，可验证次线性函数必定是凸函数；于是次线性函数本质上也将是仿射函数族的上包络．但由于次线性函数 f 还一定满足 $f(0)=0$ 等条件，我们还能得到更强的结果：

命题 2.3.1　设线性空间 X 上的函数 $f: X\to\boldsymbol{R}\cup\{+\infty\}$，且 $0\in(\mathrm{dom}\, f)^{ri}$．那么 f 是次线性函数的充要条件为：存在一族线性形式 $x_i^*\in X^*$，$i\in I$，使得

$$f(x)=\sup_{i\in I}\langle x_i^*, x\rangle,\ \forall x\in X. \tag{2.3.2}$$

证明　如果(2.3.2)成立，那么 f 显然满足(2.3.1)，因而是次线性函数．反之，令

$$L(f)=\{x^*\in X^*\mid \langle x^*, x\rangle\leqslant f(x),\ \forall x\in X\},$$

并指出当 f 是次线性函数、且满足 $0\in(\mathrm{dom}\, f)^{ri}$ 时，

$$f(x)=\sup_{x^*\in L(f)}\langle x^*, x\rangle,\ \forall x\in X. \tag{2.3.3}$$

为此，首先指出：当 $x_0\in\operatorname{dom} f$ 时，存在 $x_0^*\in L(f)$，使得

$$\langle x_0^*,\ x_0\rangle=f(x_0). \tag{2.3.4}$$

事实上，设 $L=\{x\in X \mid x=\lambda x_0,\ \lambda\in\boldsymbol{R}\}$，$x_{0L}^*\in L^*$ 定义为

$$\langle x_{0L}^*,\ \lambda x_0\rangle=\lambda f(x_0),\quad \forall\lambda\in\boldsymbol{R}. \tag{2.3.5}$$

则当 $\lambda\geqslant 0$ 时，由 f 的正齐次性，有

$$\langle x_{0L}^*,\ \lambda x_0\rangle=f(\lambda x_0).$$

当 $\lambda<0$ 时，由 f 的正齐次性和次可加性，又有

$$\langle x_{0L}^*,\ \lambda x_0\rangle=-f(-\lambda x_0)\leqslant f(\lambda x_0).$$

因此，
$$\langle x_{0L}^*,\ x\rangle\leqslant f(x),\quad \forall x\in L.$$

因为 $0\in(\operatorname{dom} f)^{ri}\cap L$，故由 Hahn-Banach 定理 2.2.6，存在 $x_0^*\in L(f)$，并满足

$$\langle x_{0L}^*,\ x\rangle=\langle x_0^*,\ x\rangle,\ \forall x\in L.$$

由(2.3.5)可知(2.3.4)成立，于是(2.3.3)对 $x\in\operatorname{dom} f$ 成立.

还需指出，当 $x_1\notin\operatorname{dom} f$ 时，对于任何 $\beta>0$，存在 $x_1^*\in L(f)$，使得 $\langle x_1^*,\ x_1\rangle\geqslant\beta$. 这时，我们必定有 $-x_1\notin\operatorname{dom} f$，否则，由 $0\in(\operatorname{dom} f)^{ri}$，存在 $\varepsilon>0$，使得

$$(-\varepsilon x_1,\ \varepsilon x_1)\subset\operatorname{dom} f.$$

由 f 的正齐次性，这与 $x_1\notin\operatorname{dom} f$ 矛盾. 这样，就有

$$f(x_1)=f(-x_1)=+\infty.$$

设
$$L_1=\{x\in X \mid x=\lambda x_1,\ \lambda\in\boldsymbol{R}\},\ x_{1L1}^*\in L_1^*$$

定义为
$$\langle x_{1L1}^*,\ \lambda x_1\rangle=\lambda\beta.$$

于是与上面同样证明，存在 $x_1^*\in L(f)$，且满足 $\langle x_1^*,\ x_1\rangle=\beta$. 这就证明了(2.3.3)对 $x\notin\operatorname{dom} f$ 也成立. ▮

推论　设 f 是满足 $0\in(\operatorname{dom} f)^{ri}$ 的次线性函数. 那么 $\operatorname{epi} f$ 是代数闭凸集. ▮

这是定理 2.2.5 与上述命题的结果.

$L(f)$实际上是 X^* 中的集合，为了更明确起见，我们把它再表示为 $K_f^*=L(f)$，即

$$K_f^*=\{x^*\in X^* \mid \langle x^*,\ x\rangle\leqslant f(x),\ \forall x\in X\}, \tag{2.3.6}$$

则已证明了当 $0\in(\operatorname{dom} f)^{ri}$ 时，

$$f(x)=\sup_{x^*\in K_f^*}\langle x^*, x\rangle.$$

K_f^* 显然是 X^* 中的代数闭凸集，而 $f(x)$ 则可称为 K_f^* 的承托函数．在 X 中则有如下的相应结果：

命题 2.3.2 设 K 为线性空间 X 中的集合，且 $K^n\neq\varnothing$，σ_K: $X^*\to \boldsymbol{R}\cup\{+\infty\}$ 是它的承托函数，即

$$\sigma_K(x^*)=\sup_{x\in K}\langle x^*, x\rangle.$$

那么

$$K=\{x\in X\,|\,\langle x^*, x\rangle\leqslant\sigma_K(x^*),\ \forall x^*\in X^*\} \qquad (2.3.7)$$

的充要条件为：K 是代数闭凸集．

证明 如果(2.3.7)成立，那么 K 显然是代数闭凸集．反之，如果 K 是代数闭凸集，$x_0\notin K$，那么由凸集强分离定理 1.4.9，存在 $x_0^*\in X^*$，使得

$$\langle x_0^*, x_0\rangle>\sup_{x\in K}\langle x_0^*, x\rangle=\sigma_K(x_0^*),$$

即 x_0 也不属于(2.3.7)的右端，从而 K 包含右端．反向关系是显然的．∎

(2.3.7)与(2.3.6)在形式上几乎一样，但由于 X 与 X^* 的地位不是平等的，故它们在含义上就不完全一致．(2.3.7)说明任何相对代数内部非空的代数闭凸集一定是一族闭半空间的交，或者说是由一族超平面所"围"成的．(2.3.6)虽然说的是在 X^* 中的类似情况，但由于当 X 为无限维时，$X^{**}\neq X$，(2.3.6)只表达了那些承托函数的有效域在 X 中的 X^* 的代数闭凸集．然而，如果 X 为有限维，这种不对称性就会消失．下一章在拓扑线性空间中讨论这一问题时，也同样可使(2.3.6)和(2.3.7)统一起来．

非负的次线性函数称为 Minkowski 函数．这种函数与包含原点的凸集紧密相关．设 $A\subset X$ 为凸集，且 $0\in A$．令

$$p_A(x)=\inf\{\alpha>0\,|\,[0, x/\alpha]\subset A\}, \qquad (2.3.8)$$

这里规定 $\inf\varnothing=+\infty$．于是有

命题 2.3.3 $p_A(x)$ 是 Minkowski 函数．

证明 $p_A(x)\geqslant 0$ 由定义(2.3.8)可得．我们验证它也满足

(2.3.1). 事实上，首先由定义，$p_A(0)=0$; 而当 $t>0$ 时，

$$p_A(tx)=\inf\{\alpha>0 \mid [0,\ tx/\alpha]\subset A\}$$
$$=t\inf\{\alpha/t>0 \mid [0,\ tx/\alpha]\subset A\}=tp_A(x),$$

故(2.3.1)的 i)成立.

另一方面，对于任何 x_1、$x_2\in X$ 和任何 $\varepsilon>0$, 存在 α_1、$\alpha_2>0$, 使得 $x_1\in\alpha_1 A$, $x_2\in\alpha_2 A$, 且

$$p_A(x_1)>\alpha_1-\varepsilon,\ p_A(x_2)>\alpha_2-\varepsilon. \tag{2.3.9}$$

由于 A 是凸集，

$$\alpha_1 A+\alpha_2 A=(\alpha_1+\alpha_2)\left(\frac{\alpha_1}{\alpha_1+\alpha_2}A+\frac{\alpha_2}{\alpha_1+\alpha_2}A\right)$$
$$=(\alpha_1+\alpha_2)A,$$

从而 $$x_1+x_2\in\alpha_1 A+\alpha_2 A=(\alpha_1+\alpha_2)A.$$

这样，由(2.3.8)、(2.3.9)，有

$$p_A(x_1+x_2)\leqslant(\alpha_1+\alpha_2)<p_A(x_1)+p_A(x_2)+2\varepsilon.$$

由于 ε 的任意性，即得 p_A 满足(2.3.1)的 ii). ▌

命题 2.3.4 设 A 为 X 中的凸集，且 $0\in A^{ri}$, p_A 如(2.3.8)所定义. 那么

$$\left.\begin{aligned}A^{ri}&=\{x\in X \mid p_A(x)<1\};\\ A^{c}&=\{x\in X \mid p_A(x)\leqslant 1\}.\end{aligned}\right\} \tag{2.3.10}$$

证明 事实上，由(1.3.5),

$$A^{ri}=\bigcup_{x\in A\setminus\{0\}}[0,\ x),$$

由(1.3.9), $$A^{c}=\{x\in X \mid [0,\ x)\subset A^{ri}\}.$$

联系(2.3.8)，易证(2.3.10)成立. 详细过程留给读者作为练习. ▌

命题 2.3.5 设 p: $X\to \boldsymbol{R}_+\cup\{+\infty\}$ 为 Minkowski 函数. 凸集 A 满足:

$$\{x\in X \mid p(x)<1\}\subset A\subset\{x\in X \mid p(x)\leqslant 1\}. \tag{2.3.11}$$

那么必定有 $p(x)=p_A(x)=\inf\{\alpha>0 \mid [0,\ x/\alpha]\subset A\}$.

证明 事实上，对于任何 $x\in X$, 有

$$p_A(x)=\inf\{\alpha>0 \mid [0,\ x/\alpha]\subset A\}$$
$$\geqslant\inf\{\alpha>0 \mid p(x/\alpha)\leqslant 1\}$$

$$=\inf\{\alpha>0\,|\,p(x)\leqslant\alpha\}=p(x);$$

而另一方面，又有

$$\begin{aligned}p(x)&=\inf\{\alpha>0\,|\,p(x)<\alpha\}\\&=\inf\{\alpha>0\,|\,p(x/\alpha)<1\}\\&\geqslant\inf\{\alpha>0\,|\,x/\alpha\in A\}=p_A(x).\end{aligned}$$

(2.3.11)

故命题得证. ▌

命题 2.3.4 把一个相对代数内部非空的凸集与一个 Minkowski 函数联系起来，且它的相对代数内部与代数闭包也都可用这个 Minkowski 函数表示. 命题 2.3.5 又说明这样的 Minkowski 函数联系的是一族有相同的相对代数内部和代数闭包的凸集. 值得注意的是：命题 2.3.5 中并无 A 的相对代数内部包含原点的要求. 于是(2.3.11)的两端又可看作相对代数内部和代数闭包概念的某种推广(这里用 A 的锥包代替 A 的仿射包来考虑).

当 $0\in(\mathrm{dom}\,p)^{ri}$ 时，Minkowski 函数 p 将联系着 X 和 X^* 的一对代数闭凸集，即

$$K_p=\{x\in X\,|\,p(x)\leqslant 1\};$$

$$K_p^*=\{x^*\in X^*\,|\,\langle x^*,\ x\rangle\leqslant p(x),\ \forall x\in X\}.$$

于是有　$\langle x^*,\ x\rangle\leqslant 1,\ \forall x^*\in K_p^*,\ \forall x\in K_p;$

且易证　$K_p=\{x\in X\,|\,\langle x^*,\ x\rangle\leqslant 1,\ \forall x^*\in K_p^*\};$

$$K_p^*=\{x^*\in X^*\,|\,\langle x^*,\ x\rangle\leqslant 1,\ \forall x\in K_p\}.$$

K_p 与 K_p^* 互称为对方的极化集. 在下章中，将对称地讨论这种对偶关系.

不取 $+\infty$ 的 Minkowski 函数称为规范(gauge)函数，这时(2.3.11)所定义的 A 是吸收凸集，即 $0\in A^i$. 通常也称对应吸收凸集 A 的 Minkowski 函数 p_A 为 A 的规范函数. 对称的规范函数，即满足

$$p(x)=p(-x),\ \forall x\in X,$$

且不取 $+\infty$ 的非负次线性函数，称为空间的半范数. 如果半范数还满足

$$p(x)=0 \Leftrightarrow x=0,$$

那么称它为空间 X 的范数. 赋有一个范数的线性空间, 当它的拓扑结构由该范数所决定时, 称为赋范空间; 赋有一族半范数的线性空间, 当它的拓扑结构由该半范数族所决定时, 称为局部凸空间. 进一步的讨论也将在下章中叙述.

§ 2.4 数学规划的 Lagrange 乘子

有了凸集和凸函数的概念以后, 作为它们的初步应用, 要在下节中讨论凸规划问题, 即求一个凸函数在一个凸集上的最小值问题. 本节中先补充一些有关数学规划论的知识.

一个数学规划问题的形式如下:

$$(\mathscr{P})\quad \begin{cases} f(x)\to \min \\ g_i(x)\leqslant 0,\ i=1,\ \cdots,\ p; \\ h_j(x)=0,\ j=1,\ \cdots,\ q. \end{cases}$$

这里 f, g_i $(i=1, \cdots, p)$, h_j $(j=1, \cdots, q)$ 都是线性空间 X 上的取广义实值的函数. 由于允许 f 取 $\pm\infty$, 上述形式的规划问题中实际上还包括把 x 限制在 X 的某个集合 S 中的问题, 因为这只需认为 f 在 S 外取 $+\infty$ 即可. 令

$$\left.\begin{aligned} K=\{x\in X \mid g_i(x)\leqslant 0,\ i=1,\ \cdots,\ p; \\ h_j(x)=0,\ j=1,\ \cdots,\ q\}; \end{aligned}\right\} \tag{2.4.1}$$

称它为问题 $(\mathscr{P})$ 的**容许集**, 其中条件

$$g_i(x)\leqslant 0\quad (i=1,\ \cdots,\ p)$$

称为问题 $(\mathscr{P})$ 的**不等式约束**, 条件

$$h_j(x)=0\quad (j=1,\ \cdots,\ q)$$

称为问题 $(\mathscr{P})$ 的**等式约束**. 如果存在 $\hat{x}\in K$, 使得

$$f(\hat{x})=\inf_{x\in K} f(x),$$

那么 $\hat{x}$ 称为问题 $(\mathscr{P})$ 的**解**; $v=\inf\limits_{x\in K} f(x)$ 则称为问题 $(\mathscr{P})$ 的值; 而 f 则称为问题 $(\mathscr{P})$ 的**目标函数**.

对于任何 $K\subset X$，可定义K的指标函数为

$$\delta_K(x)=\begin{cases}0, & \text{当 } x\in K;\\ +\infty, & \text{当 } x\notin K.\end{cases} \tag{2.4.2}$$

于是任何约束极值问题

$$\begin{cases}f(x)\to\min\\ x\in K\subset X\end{cases}$$

都可把指标函数 δ_K 当作"理想罚函数"，而化为下列无约束极值问题：

$$f(x)+\delta_K(x)\to\min.$$

对于问题($\mathscr{P}$)，其对应的K如(2.4.1)所表示．这里指出，对于这样的 K，有下列命题成立．

命题 2.4.1

$$\begin{aligned}\delta_K(x)&=\sup_{\lambda\in R_+^{p*},\mu\in R^{q*}}\{\langle\lambda,\ g(x)\rangle+\langle\mu,\ h(x)\rangle\}\\ &=\sup_{\lambda\in R_+^{p*},\mu\in R^{q*}}\left\{\sum_{i=1}^{p}\lambda_i\,g_i(x)+\sum_{j=1}^{q}\mu_jh_j(x)\right\},\end{aligned} \tag{2.4.3}$$

这里
$$g(x)=(g_1(x),\ \cdots,\ g_p(x))\in R^p;$$
$$h(x)=(h_1(x),\ \cdots,\ h_q(x))\in R^q;$$
$$\lambda=(\lambda_1,\ \cdots,\ \lambda_p)\in R_+^{p*}=\{\lambda\in R^{p*}\,|\,\lambda_i\geqslant 0,\ i=1,\ \cdots,\ p\};$$
$$\mu=(\mu_1,\ \cdots,\ \mu_q)\in R^{q*}.$$

证明 事实上，

$$\begin{aligned}&\sup_{\lambda\in R_+^{p*},\mu\in R^{q*}}\{\langle\lambda,\ g(x)\rangle+\langle\mu,\ h(x)\rangle\}\\ &=\sum_{i=1}^{p}\sup_{\lambda_i\geqslant 0}\lambda_i\,g_i(x)+\sum_{j=1}^{q}\sup_{\mu_j\in R}\mu_jh_j(x),\end{aligned}$$

而
$$\sup_{\lambda_i\geqslant 0}\lambda_ig_i(x)=\begin{cases}0, & \text{当 } g_i(x)\leqslant 0,\\ +\infty, & \text{当 } g_i(x)>0,\end{cases}\quad i=1,\ \cdots,\ p;$$

$$\sup_{\mu_j\in R}\mu_jh_j(x)=\begin{cases}0, & \text{当 } h_j(x)=0,\\ +\infty, & \text{当 } h_j(x)\neq 0,\end{cases}\quad j=1,\ \cdots,\ q.$$

由K的定义(2.4.1)和 δ_K 的定义(2.4.2)，即得(2.4.3)．∎

我们称 $L:\ X\times R_+^{p*}\times R^{q*}\to R\cup\{\pm\infty\}$，

$$L(x,\ \lambda,\ \mu)=f(x)+\langle\lambda,\ g(x)\rangle+\langle\mu,\ h(x)\rangle$$
$$=f(x)+\sum_{i=1}^{p}\lambda_i g_i(x)+\sum_{j=1}^{q}\mu_j h_j(x)$$

为问题($\mathscr{P}$)的 Lagrange 函数.

令

$$\alpha=\sup_{\lambda,\mu}\inf_{x} L(x,\ \lambda,\ \mu); \tag{2.4.4}$$

$$\beta=\inf_{x}\sup_{\lambda,\mu} L(x,\ \lambda,\ \mu). \tag{2.4.5}$$

则由命题 2.4.1 立即可得以下命题.

命题 2.4.2 问题($\mathscr{P}$)的值

$$v=\inf_{x\in K} f(x)=\beta,$$

且 $\hat{x}\in K$ 为($\mathscr{P}$)的解当且仅当

$$\sup_{\lambda,\mu} L(\hat{x},\ \lambda,\ \mu)=\beta. \ \blacksquare$$

由(2.4.4)和(2.4.5),一般只有

$$\alpha\leqslant\beta. \tag{2.4.6}$$

事实上,由 $L(x,\ \lambda,\ \mu)\leqslant\sup\limits_{\lambda,\mu} L(x,\ \lambda,\ \mu)$,

立即可得 $\inf\limits_{x\in X} L(x,\ \lambda,\ \mu)\leqslant\inf\limits_{x\in X}\sup\limits_{\lambda,\mu} L(x,\ \lambda,\ \mu)=\beta.$

因此, $\alpha=\sup\limits_{\lambda,\mu}\inf\limits_{x} L(x,\ \lambda,\ \mu)\leqslant\beta.$

然而,如果存在$(\hat{\lambda},\ \hat{\mu})\in \boldsymbol{R}_+^{p*}\times\boldsymbol{R}^{q*}$, 满足

$$\inf_{x} L(x,\ \hat{\lambda},\ \hat{\mu})=\beta=v, \tag{2.4.7}$$

那么由 $\alpha\geqslant\inf\limits_{x} L(x,\ \hat{\lambda},\ \hat{\mu})$, 就可得 $\alpha=\beta$. 满足(2.4.7)的$(\hat{\lambda},\ \hat{\mu})$称为问题($\mathscr{P}$)的 Lagrange 乘子.

根据定义, Lagrange 乘子$(\hat{\lambda},\ \hat{\mu})$必定是下列规划问题的解:

$$(\mathscr{P}^*)\begin{cases}\inf\limits_{x} L(x,\ \lambda,\ \mu)\to\max,\\ \lambda\in\boldsymbol{R}_+^{p*},\ \mu\in\boldsymbol{R}^{q*}.\end{cases}$$

称它为原问题($\mathscr{P}$)的对偶问题,因为根据命题 2.4.2, 问题($\mathscr{P}$)等价于

$$(\mathscr{P})\quad\begin{cases}\sup\limits_{\lambda,\mu} L(x,\lambda,\mu)\to\min,\\ x\in X.\end{cases}$$

问题$(\mathscr{P})$与问题$(\mathscr{P}^*)$虽然在形式上很对称，但实际上它们的地位是不平等的．这是因为 Lagrange 函数 $L(x,\lambda,\mu)$作为 x 的函数没有任何限制，而作为(λ,μ)的函数则必定是(λ,μ)的仿射函数，从而 $\inf\limits_{x} L(x,\lambda,\mu)$必定是$(\lambda,\mu)$的凹函数，也就是说，问题$(\mathscr{P}^*)$一定是一个求某凹函数在某凸集上的最大值问题，它也等价于求某凸函数在某凸集上的最小值问题．如上所述，这种问题称为凸规划问题．因此，如果期望问题$(\mathscr{P})$和$(\mathscr{P}^*)$之间有一种相互的对偶性，那么问题$(\mathscr{P})$必须是凸规划问题．我们即将看到，线性规划问题就属于这种情形；更一般的对偶理论将在第四章中讨论．

下面是数学规划理论的一条基本定理：

定理 2.4.3（鞍点定理）　下列两个命题是等价的：

i) $\hat{x}$ 是问题$(\mathscr{P})$的解，$(\hat{\lambda},\hat{\mu})$是问题$(\mathscr{P})$的 Lagrange 乘子；

ii) $(\hat{x},\hat{\lambda},\hat{\mu})$是问题$(\mathscr{P})$的 Lagrange 函数的鞍点，即

$$\begin{gathered}L(\hat{x},\lambda,\mu)\leqslant L(\hat{x},\hat{\lambda},\hat{\mu})\leqslant L(x,\hat{\lambda},\hat{\mu}),\\ \forall x\in X,\ \forall(\lambda,\mu)\in \boldsymbol{R}_+^{p^*}\times\boldsymbol{R}^{q^*}.\end{gathered}\tag{2.4.8}$$

此外，这时还有

$$\hat{\lambda}_i g_i(\hat{x})=0\quad(i=1,\cdots,p).\tag{2.4.9}$$

证明　i)⇒ii)：由命题 2.4.2 和(2.4.7)，这时有

$$\begin{gathered}L(\hat{x},\lambda,\mu)\leqslant\beta\leqslant L(x,\hat{\lambda},\hat{\mu}),\\ \forall x\in X,\ \forall(\lambda,\mu)\in\boldsymbol{R}_+^{p^*}\times\boldsymbol{R}^{q^*}.\end{gathered}$$

因此，(2.4.8)对于 $\beta=L(\hat{x},\hat{\lambda},\hat{\mu})$成立．

ii)⇒i)：由(2.4.8)有

$$\beta\leqslant\sup_{\lambda,\mu}L(\hat{x},\lambda,\mu)=L(\hat{x},\hat{\lambda},\hat{\mu})=\inf_{x}L(x,\hat{\lambda},\hat{\mu})\leqslant\alpha.$$

再由 $\alpha\leqslant\beta$ 可知

$$\sup_{\lambda,\mu}L(\hat{x},\lambda,\mu)=\inf_{x}L(x,\hat{\lambda},\hat{\mu})=L(\hat{x},\hat{\lambda},\hat{\mu})=\beta.$$

因此，由命题 2.4.2，$\hat{x}$ 是问题$(\mathscr{P})$的解；根据(2.4.7)，$(\hat{\lambda},\hat{\mu})$是

问题($\mathscr{P}$)的 Lagrange 乘子.

此外,由 $\hat{x}\in K$, 故

$$L(\hat{x},\ \hat{\lambda},\ \hat{\mu})=f(\hat{x})+\sum_{i=1}^{p}\hat{\lambda}_i g_i(\hat{x})=\beta=f(\hat{x}).$$

因此,
$$\sum_{i=1}^{p}\hat{\lambda}_i g_i(\hat{x})=0.$$

但
$$\hat{\lambda}_i\geqslant 0,\quad g_i(\hat{x})\leqslant 0,$$

以至
$$\hat{\lambda}_i g_i(\hat{x})\leqslant 0\quad (i=1,\ \cdots,\ p),$$

从而(2.4.9)成立. ▮

注 1 关系式(2.4.9)称为问题($\mathscr{P}$)的松紧关系. 它在许多实际问题中有重要意义. ▮

注 2 如果$(\hat{\lambda},\ \hat{\mu})$是问题($\mathscr{P}$)的 Lagranege 乘子, 那么问题($\mathscr{P}$)的解 $\hat{x}$ 一定也是下列问题的解:

$$(\mathscr{P}_L)\quad L(x,\ \hat{\lambda},\ \hat{\mu})\to\min.$$

这是因为(2.4.8)的右端成立. 但这并不意味着($\mathscr{P}_L$)的解也一定是($\mathscr{P}$)的解,因为($\mathscr{P}_L$)的解 $\tilde{x}$ 只能满足(2.4.8)的右端, 却不一定满足(2.4.8)的左端. 下面简单例子说明这种情况是可能的:

$$(P)\quad \begin{cases}\min x\\ |x|\leqslant 0\end{cases}$$

易证所有满足 $\hat{\lambda}\geqslant 1$ 的 $\hat{\lambda}$ 都是问题(P)的 Lagrange 乘子. 但对于 $\hat{\lambda}=1$, 所有满足 $x\leqslant 0$ 的 $\tilde{x}$ 都是问题

$$(P_L)\quad L(x,\ \hat{\lambda})=x+|x|\to\min$$

的解. 而当 $\tilde{x}<0$ 时,它并不是(P)的解. ▮

注 3 上面的结果容易推广到约束为无限维的情形. 为此, 首先需要半序线性空间的概念. 线性空间 Y 称为半序线性空间, 是指 Y 中定义了序关系"$\leqslant$", 使 Y 成为序集(参看 § 1.1), 同时满足

i) $\forall z\in Y,\ x\leqslant y\Rightarrow x+z\leqslant y+z;$

ii) $\forall\lambda\geqslant 0,\ \forall x\geqslant 0,\ \lambda x\geqslant 0.$

现在设 X、Z 是线性空间, Y 是半序线性空间,

$$f:X\to\boldsymbol{R}\cup\{\pm\infty\},\quad g:X\to Y,\quad h:X\to Z$$

为任意的映射．于是问题($\mathscr{P}$)可推广为

$$(\mathscr{P})\quad\begin{cases}f(x)\to\min,\\ g(x)\leqslant 0,\\ h(x)=0.\end{cases}$$

其 Lagrange 函数定义为

$$L:X\times Y_+^*\times Z^*\to\boldsymbol{R}\cup\{\pm\infty\},$$

$$L(x,\lambda,\mu)=f(x)+\langle\lambda,g(x)\rangle+\langle\mu,h(x)\rangle,$$

这里

$$Y_+^*=\{\lambda\in Y^*\mid\langle\lambda,y\rangle\geqslant 0,\ \forall y\geqslant 0\}.\qquad(2.4.10)$$

于是其对偶问题($\mathscr{P}^*$)以及 Lagrange 乘子等都能类似定义．这时定理 2.4.3 仍成立；松紧关系(2.4.9)则代替为

$$\langle\hat{\lambda},g(\hat{x})\rangle=0.$$

需要注意的是：(2.4.10)中的 Y_+^* 有时可能只包含唯一的零元素，这甚至在某些有意义的问题中也会发生*)．Y_+^* 包含非零元素的充分条件之一是 $Y_+=\{y\in Y\mid y\geqslant 0\}$ 有非空代数内部(参看第一章习题 11)．▌

§2.5　凸规划的 Lagrange 乘子法则

上节中讨论了一般的数学规划问题，并提出了 Lagrange 乘子的概念．由 Lagrange 乘子的定义(2.4.7)可知，它的好处在于：可把约束极值问题($\mathscr{P}$)简化为无约束极值问题(参看上节注 2)

$$(\mathscr{P}_L)\quad L(x,\hat{\lambda},\hat{\mu})\to\min.$$

因此，求出问题($\mathscr{P}$)的 Lagrange乘子对于求解($\mathscr{P}$)是有重要意义的．然而，在一般情况下，这种 Lagrange 乘子并非总是存在的，即使对于凸规划也是如此．本节要指出对于凸规划的 Lagrange

*) 例如，当 $Y=L^p[0,1]$，$0<p<1$ 时，$Y_+^*=\{0\}$，这里 $L^p[0,1]$表示$[0,1]$上的 p 次 Lebesgue 可积函数全体．参看 H. H. Schaefer: Topological Vector Spaces, Springer Verlag, New York, 1980, p. 252.

乘子的存在条件. 这类结果常称为 Lagrange 乘子法则. 为此, 先证明一条有一般意义的定理:

定理 2.5.1 设 $F_1, \cdots, F_m: X\to \boldsymbol{R}\cup\{+\infty\}$ 为线性空间 X 上的真凸函数; $h: X\to Z$ 为 X 到线性空间 Z 的仿射映射, 即

$$h(x)=L(x)+z_0,\quad \forall x\in X,$$

这里 $L: X\to Z$ 为线性映射, $z_0\in Z$. 如果

$$\max_{1\leqslant i\leqslant m} F_i(x)\geqslant 0,\quad \forall x: h(x)=0, \tag{2.5.1}$$

那么存在非零的 $(\alpha, \mu)\in \boldsymbol{R}_+^{m*}\times Z^*$, 使得

$$\begin{aligned}&\langle\alpha, F(x)\rangle+\langle\mu, h(x)\rangle\\ &=\sum_{i=1}^m \alpha_i F_i(x)+\langle\mu, h(x)\rangle\geqslant 0,\end{aligned} \tag{2.5.2}$$

$$\forall x\in\bigcap_{i=1}^m \operatorname{dom} F_i.$$

此外, 如果
$$0\in\left(h\left(\bigcap_{i=1}^m \operatorname{dom} F_i\right)\right)^i,$$

那么(2.5.2)也是(2.5.1)的充分条件, 并且这时必定有

$$\alpha=(\alpha_1, \cdots, \alpha_m)\neq 0.$$

证明 令

$$Q=\bigcap_{i=1}^m \operatorname{dom} F_i,$$

$G: Q\to\boldsymbol{R}^m\times Z$ 定义为

$$G(x)=(F_1(x), \cdots, F_m(x); h(x)).$$

又设
$$E=(\boldsymbol{R}_+^m, 0)=\{(y_1, \cdots, y_m; z)\in\boldsymbol{R}^m\times Z\mid y_i\geqslant 0,$$
$$i=1, \cdots, m; z=0\}.$$

我们指出, $G(Q)+E$ 是凸集. 事实上, 如果

$$(y; z)、(y'; z')\in G(Q)+E,$$

那么存在 x、$x'\in Q$, 使得

$$F_i(x)\leqslant y_i,\ F_i(x')\leqslant y_i'\quad (i=1, \cdots, m),$$
$$h(x)=z,\quad h(x')=z'.$$

从而对于任何 $\lambda\in[0, 1]$, 由 F_i 为真凸函数和 h 为仿射映射, 有

$$F_i((1-\lambda)x+\lambda x')\leqslant(1-\lambda)F_i(x)+\lambda F_i(x')$$
$$\leqslant(1-\lambda)y_i+\lambda y_i'\quad(i=1,\ \cdots,\ m);$$
$$h((1-\lambda)x+\lambda x')=(1-\lambda)h(x)+\lambda h(x')$$
$$=(1-\lambda)z+\lambda z'.$$

因此，　$((1-\lambda)y+\lambda y';\ (1-\lambda)z+\lambda z')\in G(Q)+E.$

另一方面，由(2.5.1)可知，$G(Q)+E$ 中不包含下列凸集中的点：

$$H=\{(y;\ z)\in \boldsymbol{R}^m\times Z\mid y_i<0,\ i=1,\ \cdots,\ m;\ z=0\}.$$

同时，由定理1.3.4和 h 的仿射性，不难验证

$$(G(Q)+E)^{ri}\neq\varnothing,\quad H^{ri}\neq\varnothing.$$

因此，由凸集分离定理1.4.8，存在

$$(\alpha,\ \mu)\in \boldsymbol{R}^{m*}\times Z^*,\ (\alpha,\ \mu)\neq 0,$$

使得　$$\langle\alpha,\ y\rangle+\langle\mu,\ z\rangle\leqslant\langle\alpha,\ y'\rangle+\langle\mu,\ z'\rangle,$$
$$\forall(y,\ z)\in H,\ \forall(y',\ z')\in G(Q)+E.$$

特别是对于任何 $y_1,\ \cdots,\ y_m<0,\ x\in Q$，有

$$\alpha_1y_1+\cdots+\alpha_my_m\leqslant\alpha_1F_1(x)+\cdots+\alpha_mF_m(x)+\langle\mu,\ h(x)\rangle.\tag{2.5.3}$$

令 $y_1,\ \cdots,\ y_m\to 0$，即得

$$\sum_{i=1}^{m}\alpha_iF_i(x)+\langle\mu,\ h(x)\rangle\geqslant 0,$$

$$\forall x\in Q=\bigcap_{i=1}^{m}\operatorname{dom}F_i$$

最后，由于(2.5.3)中，$y_1,\ \cdots,\ y_m$ 中的任何一个都可趋于 $-\infty$，故 $\alpha_1,\ \cdots,\ \alpha_m$ 中没有一个能取负值，否则，(2.5.3)不可能成立。因此，$\alpha=(\alpha_1,\ \cdots,\ \alpha_m)\in \boldsymbol{R}_+^{m*}$.

此外，当 $0\in(h(Q))^i$ 时，必定有 $\alpha=(\alpha_1,\ \cdots,\ \alpha_m)\neq 0$；否则，将有

$$\langle\mu,\ h(x)\rangle\geqslant 0,\ \forall x\in Q,$$

即　$$\langle\mu,\ z\rangle\geqslant 0,\ \forall z\in h(Q).$$

由于 $0\in(h(Q))^i$，这仅当 $\mu=0$ 时才有可能，与 $(\alpha,\ \mu)\neq 0$ 相矛盾。这样，$\alpha=(\alpha_1,\ \cdots,\ \alpha_m)\neq 0$，且

$$\alpha_1 F_1(x)+\cdots+\alpha_m F_m(x)\geqslant 0,\ \forall x\in Q; h(x)=0.$$

由此显然可导出(2.5.1). ▮

注 1 如果 $h(x)\equiv 0$, 那么定理 2.5.1 指出:

$\forall x\in X$, $\max\limits_{1\leqslant i\leqslant m} F_i(x)\geqslant 0 \Leftrightarrow \exists \alpha_i\geqslant 0$ $(i=1, \cdots, m)$不全为零

$$\forall x\in\bigcap_{i=1}^{m}\operatorname{dom} F_i,\quad \sum_{i=1}^{m}\alpha_i F_i(x)\geqslant 0. \blacksquare \tag{2.5.4}$$

注 2 值得注意的是(2.5.2)和(2.5.4)中的条件

$$\text{“}\forall x\in\bigcap_{i=1}^{m}\operatorname{dom} F_i\text{”},$$

不能改进为“$\forall x\in X$”. 下面是一个反例:

设 $X=\boldsymbol{R}$, $F_1(x)=x$, $F_2(x)=\begin{cases}-1, & \text{当 } x\geqslant 0;\\ +\infty, & \text{当 } x<0.\end{cases}$

则 F_1、F_2 都是真凸函数, 且

$$\forall x\in\boldsymbol{R},\quad \max\{F_1(x), F_2(x)\}\geqslant 0.$$

但是, 使 $\alpha_1 F_1(x)+\alpha_2 F_2(x)\geqslant 0,\ \forall x\geqslant 0,$ (2.5.5)

成立的不全为零的非负(α_1, α_2)只可能有 $\alpha_1>0$, $\alpha_2=0$, 而这样的(α_1, α_2)不可能使(2.5.5)中的“$\forall x\geqslant 0$”代替为“$\forall x\in\boldsymbol{R}$”. ▮

下面讨论凸规划的 Lagrange 乘子的存在问题. 形式为$(\mathscr{P})$的凸规划, 是指 f, g_1, $\cdots$, g_p 为 X 上的凸函数, h_1, $\cdots$, h_q 为 X 上的仿射函数, 因为这时对应的表示为(2.4.1)的容许集 K 是凸集. 对于这样一般的凸规划, Lagrange 乘子是不一定存在的(可以看到这种反例), 即使要求 f、g_i 等都是真凸函数. 在一般情况下, 只能有下列较弱的结果:

定理 2.5.2(Fritz John 条件) 设问题$(\mathscr{P})$中的 $f, g_1, \cdots, g_p$ 是线性空间 X 上的真凸函数, h_1, $\cdots$, h_q 为 X 上的仿射函数. 问题$(\mathscr{P})$的值 $V\neq\pm\infty$($V=+\infty$ 意味着 $K\cap\operatorname{dom} f=\phi$). 那么存在不全为零的 $\hat\lambda_0\geqslant 0$, $\hat\lambda_1$, $\cdots$, $\hat\lambda_p\geqslant 0$, $\hat\mu_1$, $\cdots$, $\hat\mu_q\in\boldsymbol{R}$, 使得

$$\begin{gathered}\hat\lambda_0 V\leqslant\hat\lambda_0 f(x)+\sum_{i=1}^{p}\hat\lambda_i g_i(x)+\sum_{j=1}^{q}\hat\mu_j h_j(x),\\ \forall x\in\operatorname{dom} f\cap\left(\bigcap_{i=1}^{p}\operatorname{dom} g_i\right).\end{gathered}\tag{2.5.6}$$

证明　不难验证：如果问题($\mathscr{P}$)有值 $V\neq\pm\infty$，则问题($\mathscr{P}$)等价于

$$\begin{cases}\max\limits_{1\leqslant i\leqslant p}\{f(x)-V,\ g_i(x)\}\geqslant 0,\\ \forall x;\quad h(x)=(h_1(x),\ \cdots,\ h_q(x))=0.\end{cases}$$

设 $m=1+p$，取

$$F_1(x)=f(x)-V,\quad F_{1+i}(x)=g_i(x)\quad(i=1,\ \cdots,\ p).$$

则定理 2.5.2 即归结为定理 2.5.1 的情形. ▇

如果(2.5.6)中的 $\hat{\lambda}_0>0$，则不妨假设 $\hat{\lambda}_0=1$(否则，只需两端同除以 $\hat{\lambda}_0$)；于是当所有的 g_i 都在 $\boldsymbol{R}$ 中取值时，即

$$\bigcap_{i=1}^{p}\operatorname{dom} g_i=X$$

时，(2.5.6)指出

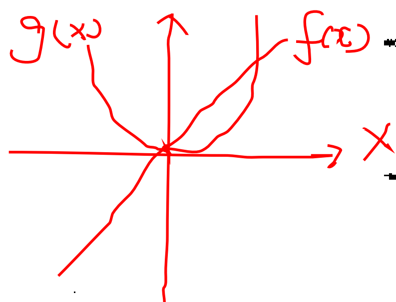
$$\beta=V\leqslant\inf_{x\in X}L(x,\ \hat{\lambda},\ \hat{\mu})\leqslant\alpha\leqslant\beta,$$

这里 α、β 如(2.4.4)、(2.4.5)所定义. 因此，

$$\inf_{x\in X}L(x,\ \hat{\lambda},\ \hat{\mu})=\alpha=\beta,$$

即 $(\hat{\lambda},\ \hat{\mu})$ 为问题($\mathscr{P}$)的 Lagrange 乘子. 但是 $\hat{\lambda}_0=0$ 是可能的. 下面是一个简单的例子：设

$$X=\boldsymbol{R},\ f(x)=x,\ g(x)=x^2.$$



则($\mathscr{P}$)为

$$\begin{cases}x\to\min,\\ x^2\leqslant 0.\end{cases}$$

它的值 V 为 0，而使

$$\hat{\lambda}_0x+\hat{\lambda}x^2\geqslant 0,\quad\forall x\in\boldsymbol{R}$$

的 $(\hat{\lambda}_0,\ \hat{\lambda})$ 只可能是 $\hat{\lambda}_0=0$，$\hat{\lambda}>0$. 这个例子说明，要保证 $\hat{\lambda}_0>0$ 还必须附加别的条件. 这种条件一般是对约束引入的，所以通常称为**约束品性条件**. 我们引入下列形式的约束品性条件：

$$(S)\begin{cases}\text{i) 存在 } x_0\in\operatorname{dom} f\text{，使得}\\ \quad g_i(x_0)<0\quad(i=1,\cdots,p);\quad h_j(x_0)=0\quad(j=1,\cdots,q).\\ \text{ii) } 0\in\left(h\left(\operatorname{dom} f\cap\left(\bigcap\limits_{i=1}^{p}\operatorname{dom} g_i\right)\right)\right)^i,\\ \quad\text{这里 } h=(h_1,\ \cdots,\ h_q).\end{cases}$$

条件(S)中的 i)称为 Slater 条件.

定理 2.5.3(Kuhn-Tucker 条件) 在定理 2.5.2 的条件下,若条件(S)成立,那么(2.5.6)中的 $\hat{\lambda}_0=1$,即存在 $\hat{\lambda}_1, \cdots, \hat{\lambda}_p \geqslant 0$, $\hat{\mu}_1, \cdots, \hat{\mu}_q \in \boldsymbol{R}$,使得

$$V \leqslant f(x)+\sum_{i=1}^{p} \hat{\lambda}_i g_i(x)+\sum_{j=1}^{q} \hat{\mu}_j h_j(x), \quad \forall x \in \operatorname{dom} f \cap\left(\bigcap_{i=1}^{p} \operatorname{dom} g_i\right). \tag{2.5.7}$$

证明 如果 $\hat{\lambda}_0=0$, 那么由(2.5.6),

$$\sum_{i=1}^{p} \hat{\lambda}_i g_i(x)+\sum_{j=1}^{q} \hat{\mu}_j h_j(x) \geqslant 0, \quad \forall x \in \operatorname{dom} f \cap\left(\bigcap_{i=1}^{p} \operatorname{dom} g_i\right). \tag{2.5.8}$$

由条件(S)中的 i),在上式中取 $x=x_0$, 则得

$$\sum_{i=1}^{p} \hat{\lambda}_i g_i(x_0) \geqslant 0.$$

但 $\hat{\lambda}_i \geqslant 0$, $g_i(x_0)<0$ $(i=1, \cdots, p)$, 上式仅当 $\hat{\lambda}_1, \cdots, \hat{\lambda}_p=0$ 时才有可能. 这样,再由(2.5.6)可得

$$\sum_{j=1}^{q} \hat{\mu}_j h_j(x)=\langle\hat{\mu}, h(x)\rangle \geqslant 0,$$

$$\forall x \in \operatorname{dom} f \cap\left(\bigcap_{i=1}^{p} \operatorname{dom} g_i\right).$$

由条件(S)中的 ii), 上式仅当

$$\hat{\mu}=(\hat{\mu}_1, \cdots, \hat{\mu}_q)=0$$

时才有可能, 这与

$$\hat{\lambda}_0, \hat{\lambda}_1, \cdots, \hat{\lambda}_p, \hat{\mu}_1, \cdots, \hat{\mu}_q$$

不全为零相矛盾. 因此, $\hat{\lambda}_0>0$, 特别是可取 $\hat{\lambda}_0=1$. ▌

推论 在定理 2.5.2 的条件下,如果条件(S)成立,且 $g_1, \cdots, g_p$ 都在 $\boldsymbol{R}$ 中取值, 那么问题$(\mathscr{P})$的 Lagrange 乘子存在; 特别是对偶问题$(\mathscr{P}^*)$的解存在. ▌

注 1 根据定理 2.5.1 的注 2, 我们不能由(2.5.7)来断定问题$(\mathscr{P})$有 Lagrange 乘子. ▌

注 2　条件(S)中的 ii)可减弱为

$$0\in\left(h\left(\operatorname{dom} f\cap\left(\bigcap_{i=1}^{p}\operatorname{dom} g_i\right)\right)\right)^{ri}.$$

因为这时我们可以把 $\boldsymbol{R}^q$ 代替为

$$\operatorname{aff} h\left(\operatorname{dom} f\cap\left(\bigcap_{i=1}^{p}\operatorname{dom} g_i\right)\right)$$

$$=\operatorname{lin} h\left(\operatorname{dom} f\cap\left(\bigcap_{i=1}^{p}\operatorname{dom} g_i\right)\right)$$

来进行同样的讨论. ▌

注 3　如果把(S)的 ii)中的 i 减弱为 ri，那么下列条件是使(S)成立的充分条件：

(S')　存在 $x_0\in(\operatorname{dom} f)^i\cap\left(\bigcap_{i=1}^{p}(\operatorname{dom} g_i)^i\right)$，使得

$g_i(x_0)<0\ (i=1,\ \cdots,\ p)$；$h_j(x_0)=0\ (j=1,\ \cdots,\ q)$.

因为仿射映射总是把集合的代数内点变为相对代数内点. ▌

注 4　等式约束显然可推广到无限维情形. 这时定理 2.5.2 和 2.5.3 都几乎不必作修改. 如果还希望把不等式约束也推广到无限维的情形，那么首先要把凸函数的概念推广为取半序线性空间中的值的凸映射. 其次，定理 2.5.2 的证明不能再由定理 2.5.1 得到，而必须重新证明. 其证明思路仍可仿照定理 2.5.1 的证明，但为了能应用凸集分离定理，需假设半序线性空间的正锥

$$Y_+=\{y\in Y\mid y\geqslant 0\}$$

有非空代数内部. 它同时也是定义条件(S)中的 $g(x)<0$ 所必须的. ▌

注 5　当 $\operatorname{dom} f\cap\left(\bigcap_{i=1}^{p}\operatorname{dom} g_i\right)=X$ 时，条件(S)的 ii)对于相对代数内部总是满足的. 这时条件(S)可减弱为：

(S'')　对于任何非仿射的 g_i，存在 $x_i\in X$，使得

$g_i(x_i)<0,\ g_k(x_i)\leqslant 0\ (k=1,\ \cdots,\ p)$,

$h_j(x_i)=0\ (j=1,\ \cdots,\ q)$.

事实上，这时不妨设所有满足(S'')的 g_i 为 $g_1,\ \cdots,\ g_{p'}\ (p'\leqslant p)$，而 $g_{p'+1},\ \cdots,\ g_p$ 为仿射函数. 还不妨设 $h_1,\ \cdots,\ h_q$ 线性无关(任何

一个 $h_{j'}$ 不能表示为其他的 h_j 的线性组合)，否则，可取出其中最大的线性无关集来作讨论，因为其他的等式约束实际上是多余的，其对应的 $\hat{\mu}_j$ 可取为零．同样还不妨假设 $g_{p'+1}, \cdots, g_p, h_1, \cdots, h_q$ 也线性无关，否则，同样可去掉一些多余的约束，并令其对应的 $\hat{\lambda}_i$、$\hat{\mu}_j$ 为零．这样一来，由(2.5.8)和(S'')，我们可得

$$\hat{\lambda}_i g_i(x_i)+\sum_{k\neq i}\lambda_k g_k(x_i)\geqslant 0 \quad (i=1, \cdots, p'),$$

从而由 $\hat{\lambda}_i\geqslant 0$, $g_i(x_i)<0$ 可得

$$\hat{\lambda}_i=0 \ (i=1, \cdots, p').$$

另一方面，又有

$$\sum_{i=p'+1}^{p}\hat{\lambda}_i g_i(x)+\sum_{j=1}^{q}\hat{\mu}_j h_j(x)=0, \ \forall x\in X.$$

由假设，这仅当 $\hat{\lambda}_{p'+1}, \cdots, \hat{\lambda}_p, \hat{\mu}_1, \cdots, \hat{\mu}_q$ 全为零时才有可能，同样导致矛盾．注意到这点是有用的，它说明对于只有仿射约束的线性规划(见下节)来说，总能有 $\lambda_0=1$. ▮

§2.6 线性规划和 Lagrange 乘子的经济解释

下列类型的极值问题称为线性规划问题：

$$(\mathscr{L})\begin{cases}\sum_{i=1}^{n} b_i x^i \to \min, \\ x^i\geqslant 0, \ i=1, \cdots, n; \\ \sum_{i=1}^{n} a_i^j x^i\geqslant c^j, \ j=1, \cdots, p.\end{cases}$$

这里 x^i, b_i, a_i^j, c^j 等都是实数．它也可以简记为

$$(\mathscr{L})\begin{cases}\langle b, x\rangle_n \to \min, \\ x\geqslant 0; \\ Ax\geqslant c,\end{cases}$$

这里 $x\in \boldsymbol{R}^n$, $b\in \boldsymbol{R}^{n*}$, $c\in \boldsymbol{R}^p$, $x\geqslant 0$, $\Leftrightarrow x^i\geqslant 0$ $(i=1, \cdots, n)$, $A: \boldsymbol{R}^n\to \boldsymbol{R}^p$ 为线性映射，它可用矩阵

$$(a_i^j) \ (i=1, \cdots, n; \ j=1, \cdots, p)$$

来表示，$Ax\geqslant c$ 的定义类似.

如所周知，有许许多多实际问题可归结为线性规划问题. 一种典型的经济解释是这样的：$x=(x^1, \cdots, x^n)$ 代表生产过程中所需要的 n 种原料的投入量，简称为投入丛，$b=(b_1, \cdots, b_n)$ 代表这 n 种原料的单位价格，简称为投入价格系，于是

$$\langle b, x\rangle_n=\sum_{i=1}^n b_i x^i$$

就是所有投入量的总价值，即生产的成本，$c=(c^1, \cdots, c^p)$ 代表 p 种产品的产出量，简称为产出丛，矩阵 A 则代表用 n 种投入原料来生产 p 种产出产品的消耗系数，它称为投入产出矩阵. 于是 $Ax\geqslant c$ 意味着要求投入丛 x 能用来生产不比 c 更少的产出丛，而问题 $(\mathscr{L})$ 就是在既定的产出目标下，要求投入的成本最小.

对应问题 $(\mathscr{L})$ 的 Lagrange 函数为

$$\begin{aligned}L(x, \nu, \lambda)&=\langle b, x\rangle_n-\langle \nu, x\rangle_n+\langle \lambda, c-Ax\rangle_p\\&=\sum_{i=1}^n\Big(b_i-\nu_i-\sum_{j=1}^p a_i^j\lambda_j\Big)x^i+\sum_{j=1}^p\lambda_j c^j.\end{aligned}$$

从而其 Lagrange 乘子 $(\hat{\nu}, \hat{\lambda})$ 首先应是使

$$\inf_{x\in R^n} L(x, \nu, \lambda)\to\max$$

的解. 但由上式可知

$$\inf_{x\in R^n} L(x, \nu, \lambda)=\begin{cases}\sum_{j=1}^p\lambda_j c^j=\langle\lambda, c\rangle_p, & \text{当 } \nu=b-A^*\lambda\geqslant 0;\\ -\infty, & \text{其他.}\end{cases}$$

这里 A^* 表示 A 的转置，即

$$A^*\lambda=\Big(\sum_{j=1}^p a_i^j\lambda_j\Big)\in \boldsymbol{R}^{n*}.$$

因此，$\hat{\lambda}$ 将是下列问题的解：

$$(\mathscr{L}^*)\quad\begin{cases}\sum_{j=1}^p\lambda_j c^j\to\max\\ \lambda_j\geqslant 0,\ j=1, \cdots, p;\\ \sum_{j=1}^p a_i^j\lambda_j\leqslant b_i,\ i=1, \cdots, n.\end{cases}$$

或

$$(\mathscr{L}^*)\begin{cases}\langle\lambda,\ c\rangle_p\to\max\\ \lambda\geqslant 0;\\ A^*\lambda\leqslant b.\end{cases}$$

$\hat{v}$ 由 $\hat{v}=b-A^*\hat{\lambda}$ 可得. $(\mathscr{L}^*)$称为$(\mathscr{L})$的对偶问题. 它也是个线性规划问题,并且通过改变符号求$(\mathscr{L}^*)$的对偶问题,可得

$$(\mathscr{L}^{**})=(\mathscr{L}).$$

因此,$(\mathscr{L})$和$(\mathscr{L}^*)$可称为互为对偶的线性规划问题.

在上面提到的$(\mathscr{L})$的经济解释下, λ 可解释为产出价格系,从而

$$\langle\lambda,\ c\rangle_p=\sum_{j=1}^{p}\lambda_j c^j$$

就是产出为 c 时的总价值, 即生产的收入, $(\mathscr{L}^*)$则可解释为选取适当的产出价格系,使生产的收入最大,条件是其价格水平不超过投入的价格水平,即对任何投入丛 x 和产出丛 $y=Ax$, λ 的选择总满足利润(收入与成本的差)非正的要求:

$$\begin{aligned}\langle\lambda,\ y\rangle_p-\langle b,\ x\rangle_n&=\langle\lambda,\ Ax\rangle_p-\langle b,\ x\rangle_n\\ &=\langle A^*\lambda-b,\ x\rangle_n\leqslant 0,\ \forall x\in\boldsymbol{R}^n.\end{aligned}$$

这样一来,也就是说,产出水平一定的成本最小问题与价格水平一定的收入最大问题是互为对偶问题.

令$(\mathscr{L})$和$(\mathscr{L}^*)$的容许集分别为

$$K_{\mathscr{L}}=\{x\in\boldsymbol{R}^n\mid x\geqslant 0,\ Ax\geqslant c\};$$

$$K_{\mathscr{L}}^*=\{\lambda\in\boldsymbol{R}^{p*}\mid\lambda\geqslant 0,\ A^*\lambda\leqslant b\}.$$

根据前面的结果,即可以得到

定理 2.6.1 下列命题是等价的:

a) $(\mathscr{L})$有解 $\hat{x}$;

a′) $(\mathscr{L})$有有限值;

b) $(\mathscr{L}^*)$有解$\hat{\lambda}$;

b′) $(\mathscr{L}^*)$有有限值;

c) $(\mathscr{L})$和$(\mathscr{L}^*)$的容许集 $K_{\mathscr{L}}$, $K_{\mathscr{L}^*}$ 都是非空的.

这时,有

$$\inf_{x\in K_{\mathscr{L}}}\langle b,\ x\rangle_n=\sup_{\lambda\in K_{\mathscr{L}^*}}\langle\lambda,\ c\rangle_p,$$

且有下列松紧关系：

$$\langle\hat{\lambda},\ A\hat{x}-c\rangle_p=\langle b-A^*\hat{\lambda},\ \hat{x}\rangle_n=0,$$

或

$$\begin{aligned}&\hat{\lambda}_j\left(\sum_{i=1}^{n}a_i^j\hat{x}^i-c^j\right)=0,\quad j=1,\ \cdots,\ p;\\&\left(b_i-\sum_{j=1}^{p}a_i^j\hat{\lambda}_j\right)\hat{x}^i=0,\quad i=1,\ \cdots,\ n.\end{aligned}\tag{2.6.1}$$

证明 事实上，由定理 2.5.3 的推论和注 4，可得 a′)⇒b)和 b′)⇒a)．而 a)⇒a′)和 b)⇒b′)是显然的．a)+b)⇒c)也是显然的．反之，由 c)可得

$$\inf_{x\in K_{\mathscr{L}}}\langle b,\ x\rangle_n=\inf_x\sup_{\nu,\lambda}L(x,\ \nu,\ \lambda)=\beta<+\infty;$$

$$\sup_{\lambda\in K_{\mathscr{L}}}\langle\lambda,\ c\rangle_p=\sup_{\nu,\lambda}\inf_x L(x,\ \nu,\ \lambda)=\alpha>-\infty.$$

但 $\beta\geqslant\alpha$ 总是成立的，因此，

$$-\infty<\sup_{\lambda\in K_{\mathscr{L}}}\langle\lambda,\ c\rangle_p\leqslant\inf_{x\in K_{\mathscr{L}}}\langle b,\ x\rangle_n<+\infty,$$

即 a′和 b′)成立．其他结论都可由定理 2.5.3 和 2.4.3 得到．∎

联系着上面的经济解释，松紧关系将有如下的意义：如果

$$\sum_{i=1}^{n}a_i^j\hat{x}^i-c^j>0,$$

即在成本最小的前提下，预定的第 j 种产出指标 c^j 居然还能超额完成，那么必须有 $\hat{\lambda}_j=0$，即这种产出的价格必定为零，它的超额完成并不能使收入增加；如果

$$b_i-\sum_{j=1}^{p}a_i^j\hat{\lambda}_j>0,$$

即在收入最大的前提下，第 i 种投入的价格 b_i 居然超过它所创收入的价格（称为"影子价格"），那么必须有 $\hat{x}^i=0$，即这种投入的投入量必定为零，也即不应该使用这种价格昂贵的原料，以免增加成本．Lagrange 函数与 Lagrange 乘子现在也有以下有趣的解释：设问题($\mathscr{L}$)的 Lagrange 乘子为 $(\hat{\nu},\ \hat{\lambda})=(b-A^*\hat{\lambda},\ \hat{\lambda})$．则问题($\mathscr{L}$)的解也是下列问题的解：

$$L(x, \hat{v}, \hat{\lambda})=\langle b, x\rangle_n-\langle b-A^*\hat{\lambda}, x\rangle_n+\langle\hat{\lambda}, c-Ax\rangle_p$$
$$=\langle A^*\lambda, x\rangle_n+\langle\hat{\lambda}, c-Ax\rangle_p\to\min.$$

这里后一项可解释为完不成生产指标所造成的损失；前一项则可解释为在影子价格系 $A^*\hat{\lambda}$ 下的成本(称为机会成本)．这就是说，在一定的产出指标要求下的成本最小的问题等价于：没有产出指标要求，甚至允许不消耗原料，而买进原料($x_i<0$)时，机会成本和完不成产出指标的损失之和最小的问题．这些结论都是相当深刻的．

对于一般的问题($\mathscr{P}$)，其 Lagrange 函数和 Lagrange 乘子都可以有类似的解释，即 Lagrange 函数可解释为要求最小的目标函数与破坏约束的惩罚函数的和，而 Lagrange 乘子则可解释为某种意义下的"最优"单位惩罚．为了更清楚地说明这一点，再举一个较具体的例子如下：

设某计划部门有一笔资金 a，将在 n 个企业之间进行分配．如果第 i 个企业分配到的资金数为 x_i，那么它将得到收益为 $R_i(x_i)$．我们假定每一企业对资金的利用率都有某种饱和趋势，或至多只能使收益按比例增长，于是可假设 $R_i(x_i)$ 都是 x_i 的凹函数．这样，为使分配达到最优，对于计划部门来说，它就面临下列规划问题：

$$\begin{cases}\sum_{i=1}^{n} R_i(x_i)\to\max;\\ x_1+\cdots+x_n\leqslant a;\\ x_1, \cdots, x_n\geqslant 0.\end{cases}\tag{2.6.2}$$

当企业个数很多时，一方面由于计划部门很难确切掌握每一 $R_i(x_i)$，以至形不成问题；另一方面即使都掌握了，这也是个计算量很大的问题，求解有困难．

于是自然会提出这样的问题：能否采取适当的办法，不是由计划部门全盘决策，而是发挥各企业的主观能动性，给各企业一定的自主权，通过由它们自行提出申请贷款、并支付利息的办法来分配这一笔资金？这样，设贷款的利率为 $\hat{\lambda}$，问题就变为一大堆由各单位自行决策的问题：

$$\begin{cases} R_i(x_i)-\hat{\lambda}x_i \to \max, \\ x_i \geqslant 0, \end{cases} \quad (i=1, \cdots, n). \qquad (2.6.3)$$

这些问题的个数虽多，但都是单变量规划问题，且每个问题都由一个企业来解，因此，实际上问题已大大简化．现在的新问题在于：提出怎样的利率 $\hat{\lambda}$，使得这笔资金仍能达到最优分配，即仍能达到总收益最大这一目标．从直观上可以看出，如果 $\hat{\lambda}$ 定得过高，那么各企业都不大愿意贷款，从而这笔资金 a 得不到充分利用；但如果 $\hat{\lambda}$ 定得过低，又会使各企业要求贷款过多，以致总数会超过资金总量 a．那么怎样是恰到好处的利率 $\hat{\lambda}$ 呢？稍加分析，我们就会发现它恰好就是对应不等式约束 $\sum\limits_{i=1}^{n} x_i \leqslant a$ 的 Lagrange 乘子．

事实上，撇开一些简单的变换就可看出，如果 $\hat{\lambda}$ 是对应 $\sum\limits_{i=1}^{n} x_i \leqslant a$ 的 Lagrange 乘子，那么问题(2.6.2)等价于

$$\begin{cases} \sum\limits_{i=1}^{n} R_i(x_i)-\hat{\lambda}\left(\sum\limits_{i=1}^{n} x_i - a\right) \to \max, \\ x_i \geqslant 0 \quad (i=1, \cdots, n). \end{cases} \qquad (2.6.4)$$

而(2.6.4)与(2.6.3)显然是等价的．$\hat{\lambda}$ 在这里恰好起着对破坏约束 $\sum\limits_{i=1}^{n} x_i \leqslant a$ 应付出的单位代价的作用．为使这种惩罚达到极大，$\hat{\lambda}$ 应该是下列问题的解：

$$\begin{cases} \sup\limits_{x\in R_+^n}\left\{\sum\limits_{i=1}^{n} R_i(x_i)-\lambda\left(\sum\limits_{i=1}^{n} x_i - a\right)\right\} \to \min, \\ \lambda \geqslant 0. \end{cases} \qquad (2.6.5)$$

由于约束函数是仿射函数，又假定所有 $R_i(x_i)$ 都是凹函数，这样的 $\hat{\lambda}$ 必定存在．问题(2.6.5)一方面是单变量规划问题，计算较简单；另一方面，我们以后还会指出，计算 $\hat{\lambda}$ 只需知道最大总收益

$$R=\sup \sum_{i=1}^{n} R_i(x_i)$$

与资金总量 $a=\sum\limits_{i=1}^{n} x_i$ 的关系．因此，在掌握情况的要求上也比原问题(2.6.2)的要求低．

问题(2.6.2)变为问题(2.6.3)的过程称为分散化，Lagrange乘子 $\hat{\lambda}$ 在这里称为分散化参数. 不但如此，上述讨论很容易推广到 a 以及 x_i 等都是向量的情形，即 a 可代表一系列不同的物资，而问题(2.6.2)则变为计划部门对这些物资的最优分配问题. 这时对应的 Lagrange 乘子 $\hat{\lambda}=(\hat{\lambda}_1, \cdots, \hat{\lambda}_p)$ 的每个分量可代表每种物资的某种价格，它们也称为影子价格. 这种利用 Lagrange 乘子的分散化方法在许多经济问题中都能得到实际实用.

第二章习题

1. 试指出，引理 2.1.3 中的条件可减弱为"除了可数个点外，f 处处有右导数 f'_+"，从而定理 2.1.4 中的条件也可作同样的减弱.
2. 设 f 是 $(0, +\infty)$ 上的递增凸函数. 试证明，或者 f 为常数，或者
$$\lim_{x\to+\infty} f(x)=+\infty.$$
3. 设 f 是 $(0, +\infty)$ 上的凸函数. 试证明 $\lim\limits_{x\to+\infty} f(x)/x$ 或存在或为 $+\infty$，并且这个极限也等于
$$\lim_{x\to\infty} f'_+(x) \quad 和 \quad \lim_{x\to\infty} f'_-(x).$$
4. 设 f 为 (a, b) 上的连续函数. 试证明 f 是 (a, b) 上的凸函数的充要条件为：对于任何 $x\in(a, b)$，有
$$\limsup_{h\to 0} \frac{f(x+h)+f(x-h)-2f(x)}{h^2}\geqslant 0.$$
5. 设 f 为 (a, b) 上的函数，且 $(a, b)\subset(0, \infty)$. 试指出，$f(1/x)$ 是 (a, b) 上的凸函数的充要条件为：$xf(x)$ 是 (a, b) 上的凸函数.
6. 试指出，凸函数的定义(2.2.1)可代替为：
$$\forall x_1、x_2\in K,\ \forall\lambda\notin[0, 1]:\ (1-\lambda)x_1+\lambda x_2\in K,$$
$$f((1-\lambda)x_1+\lambda x_2)\geqslant(1-\lambda)f(x_1)+\lambda f(x_2).$$
7. 设 $f: K\to \boldsymbol{R}$ 为凸集 K 上的凸函数，且对于任何 $x\in K$，$f(x)\geqslant 0$. 试指出，$f^2(x)$ 也是 K 上的凸函数. 试用 $\boldsymbol{R}$ 上的凸函数为例说明，条件 $f(x)\geqslant 0$ 不能去掉.
8. 线性空间 X 上的广义实值函数 $f: X\to \boldsymbol{R}\cup\{\pm\infty\}$ 称为拟凸函数，是指
$$\forall x_1、x_2\in X,\ \forall\lambda\in[0, 1],$$
$$f((1-\lambda)x_1+\lambda x_2)\leqslant\max\{f(x_1), f(x_2)\}.$$
试证明，f 拟凸的充要条件为对于任何 $\alpha\in\boldsymbol{R}$

$$\{x\in X \mid f(x)\leqslant\alpha\}$$

为凸集.

9. 设 f、g 为线性空间 X 上的凸函数. 试证明

$$F(x)=\inf_{y\in X}\{f(y)+g(x-y)\}$$

也是 X 上的凸函数.

10. 设 f 为 X 上的凸函数. 如果存在 $x_0\in(\operatorname{dom} f)^i$ 使得

$$f(x_0)=\sup_{x\in X}f(x),$$

试证明, f 恒为常数.

11. 设 f 为 (a, b) 上的连续函数. 试证明, f 是凸函数的充要条件为

$$f\left(\frac{x_1+x_2}{2}\right)\leqslant\frac{1}{2}(f(x_1)+f(x_2)),\ \forall x_1、x_2\in(a, b).$$

试指出, f 连续的条件不能去掉, 但可减弱为在 (a, b) 中的一点邻域中上有界.

[提示: 把 $\boldsymbol{R}$ 看作有理数域 $\boldsymbol{Q}$ 上的线性空间, 则 $\boldsymbol{R}$ 也有 Hamo 基, 使得每个 $x\in\boldsymbol{R}$ 都可用它们作有理线性表示. 由此可构造出关于有理数线性的 $\boldsymbol{R}$ 上的不连续函数.]

12. 试用反例指出, Hahn-Banach 定理 2.2.6 中的条件 $(\operatorname{dom} f)^{ri}\cap L\neq\varnothing$ 不能去掉. 再指出该条件可减弱为 $0\in(\operatorname{dom} f-L)^{ri}$.

13. 试用反例指出, 命题 2.3.1 中的条件 $0\in(\operatorname{dom} f)^{ri}$ 也不能去掉. [提示: 对定理 1.3.12 中的 A_0 作 p_{A_0} 这个 A_0 因 $A_0^i=\varnothing$ 也不满足命题 2.3.2.]

14. 设 f 为 X 上的正齐次函数, 即对于任何 $\lambda\geqslant 0$, 有

$$f(\lambda x)=\lambda f(x).$$

试证明, f 为凸函数当且仅当

$$f(x+y)\leqslant f(x)+f(y),\quad \forall x、y\in X.$$

15. 设 $A_1, \cdots, A_n$ 为包含原点的凸集, $p_{A_1}, \cdots, p_{A_n}$ 如 (2.3.8) 所定义. 试用 $p_{A_1}, \cdots, p_{A_n}$ 表示 $p_{\bigcap\limits_{i=1}^{n}A_i}$.

16. 试指出, 在 (2.5.4) 中关于 $F_i(x)$ 的两个不等式改为严格不等号, 则结果仍成立.

17. 设 $x_1^*, \cdots, x_m^*\in X^*$. 试证明, 不等式组

$$\langle x_i^*, x\rangle<0\quad(i=1, \cdots, m)$$

有解的充要条件为不存在不全为零的 $\alpha_1, \cdots, \alpha_m\geqslant 0$, 使得

$$\sum_{i=1}^{m}\alpha_i x_i^*=0.$$

18. 设 $x_0^*, x_1^*, \cdots, x_m^*\in X^*$, 试证明, 不等式组

$$\langle x_i^*,\ x\rangle\leqslant 0\quad(i=1,\ \cdots,\ m),\ \langle x_0^*,\ x\rangle<0.$$

有解的充要条件为不存在 $\alpha_1,\ \cdots,\ \alpha_m\geqslant 0$，使得

$$x_0^*+\sum_{i=1}^m \alpha_i x_i^*=0.$$

19. 设 $x_1^*,\ \cdots,\ x_m^*\in X^*$，$a_1,\ \cdots,\ a_m\in \boldsymbol{R}$. 试证明，不等式组

$$\langle x_i^*,\ x\rangle-a_i<0\quad(i=1,\ \cdots,\ m)$$

有解的充要条件为不存在不全为零的 $\alpha_1,\ \cdots,\ \alpha_m\geqslant 0$，使得

$$\sum_{i=1}^m a_i x_i^*=0,\quad \sum_{i=1}^m \lambda_i a_i\leqslant 0.$$

[樊畿(Ky Fan)定理].

20. 试用例子说明，Slater 条件(S)不是凸规划的 Lagrange 乘子存在的必要条件.

21. 试把线性规划对偶性定理 2.6.1 推广到无限维情形.

22. 试把线性规划对偶问题($\mathscr{L}$)和($\mathscr{L}^*$)作另一种经济解释：λ：产出丛；b：投入丛；c：产出价格系；x：投入(影子)价格系；A：产出投入矩阵；并讨论有关结论的经济含义.

23. 下列形式的规划问题称为二次规划：

$$\begin{cases}\left\{\dfrac{1}{2}\langle Ax,\ x\rangle+\langle b,\ x\rangle\right\}\to \min,\\ x\geqslant 0,\\ Bx\geqslant c.\end{cases}$$

这里 $x\in\boldsymbol{R}^n$，A：$\boldsymbol{R}^n\to\boldsymbol{R}^{n*}$ 为对称正定($n\times n$)矩阵，$c\in\boldsymbol{R}^p$，B：$\boldsymbol{R}^n\to\boldsymbol{R}^p$ 为线性映射($n\times p$ 矩阵). 试讨论它的对偶问题和 Lagrange 乘子.

第三章　拓扑线性空间中的凸性

§3.1　拓扑空间及其有关概念

在前两章中讨论了线性空间中的凸集和凸函数以及它们的应用．由于线性空间仅仅是一种代数结构，因而牵涉到极限概念的问题(例如凸函数的连续性)就不便在线性空间这一框架中讨论．为此，我们的框架还需要一种拓扑结构．下面首先简要地补充一些有关拓扑空间的知识．

设 X 为一个集合，$\mathscr{O}$ 为 X 的一个子集族，且满足：

O_1: $X\in\mathscr{O}$，$\phi\in\mathscr{O}$；

O_2: 如果 $O_i\in\mathscr{O}$，$i\in I$，那么 $\bigcup\limits_{i\in I}O_i\in\mathscr{O}$；

O_3: 如果 $O_i\in\mathscr{O}$ $(i=1,\cdots,n)$，那么

$$\bigcap_{i=1}^{n}O_i\in\mathscr{O}.$$

则称 X 是以 $\mathscr{O}$ 为开集族的拓扑空间，$\mathscr{O}$ 中的元素称为 X 的开集．例如，n 维空间对于通常意义下的开集族就构成拓扑空间．一个线性空间如果以代数开集为开集，也构成一个拓扑空间(参看命题 1.3.1 的注 1)．

同一个集合 X 中可能用许多种不同的方式来定义它的开集族，或者说 X 中可能定义许多种不同的拓扑．如果 X 上定义了两种开集族 $\mathscr{O}_1$ 和 $\mathscr{O}_2$，且 $\mathscr{O}_1\subset\mathscr{O}_2$，则称用 $\mathscr{O}_2$ 定义的拓扑比用 $\mathscr{O}_1$ 定义的拓扑要强或细；反之，则称为弱或粗．任何两种拓扑当然不一定都能比较强弱、粗细．但是在所有可能的拓扑中都有一种最细的拓扑和最粗的拓扑．最细的拓扑即取 $\mathscr{O}$ 为 X 的所有子集全体，这种拓扑称为 X 的离散拓扑；这时 X 中的每个单点集都是开集．最粗的拓扑即取 $\mathscr{O}$ 只有两个元素：X 和 $\varnothing$，这种拓扑称为 X

的平凡拓扑.

如果 $O\in\mathscr{O}$ 为 X 的开集, 那么 $F=X\setminus O$ 称为 X 的闭集. 闭集的全体组成拓扑空间 X 的闭集族 $\mathscr{F}$, 它满足:

F_1: $\phi\in\mathscr{F}$; $X\in\mathscr{F}$;

F_2: 如果 $F_i\in\mathscr{F}$, $i\in I$, 那么 $\bigcap_{i\in I}F_i\in\mathscr{F}$;

F_3: 如果 $F_i\in\mathscr{F}$ $(i=1, \cdots, n)$, 那么

$$\bigcup_{i=1}^{n}F_i\in\mathscr{F}.$$

这是由下列集合运算法则所导得的:

$$X\setminus\bigcup_{i\in I}O_i=\bigcap_{i\in I}(X\setminus O_i),$$

$$X\setminus\bigcap_{i=1}^{n}O_i=\bigcup_{i=1}^{n}(X\setminus O_i).$$

拓扑空间也可通过给定闭集族来定义, 即先引入闭集族, 然后, 令闭集的余集为开集.

由开集和闭集的概念出发, 可定义出许多别的拓扑概念来, 设 $A\subset X$ 为拓扑空间 X 的子集. 那么 A 的所有开子集的并集称为 A 的内部, 记作 $\operatorname{int}A$; 所有含 A 的闭集的交集称为 A 的闭包, 记作 $\overline{A}$. $\operatorname{int}A$ 中的点称为 A 的内点, $\overline{A}$ 中的点称为 A 的接触点, $\overline{A}\setminus\operatorname{int}A$ 中的点则称为 A 的边界点. 如果 $A\subset B$, 但 $B\subset\overline{A}$, 则称 A 在 B 中稠密. 如果 $\overline{A}=X$, 则称 A 是 X 中的稠密集. 当 X 有可数的稠密集时, 则称 X 是可分的.

一个更重要的拓扑概念是邻域的概念. 设 $x\in X$, $V\subset X$. 如果 $x\in\operatorname{int}V$, 那么 V 称为点 x 的邻域. 设 $\mathscr{V}(x)$ 表示 x 的邻域的全体. 那么 $\mathscr{V}(x)$ 满足:

V_1: $\forall V\in\mathscr{V}(x)$, $x\in V$;

V_2: 如果 U、$V\in\mathscr{V}(x)$, 那么 $U\cap V\in\mathscr{V}(x)$;

V_3: 如果 $U\in\mathscr{V}(x)$, $U\subset V$, 那么 $V\in\mathscr{V}(x)$;

V_4: 如果 $U\in\mathscr{V}(x)$, 那么存在 $V\in\mathscr{V}(x)$, 使得

$$\forall y\in V,\quad U\in\mathscr{V}(y).$$

这里 V_1、V_2、V_3 都是显然的, 而 V_4 中只需取 $V=\operatorname{int}U$ 即可. 另

一方面，拓扑空间也可通过邻域来定义，即对每一 $x\in X$，定义出满足 V_1，V_2，V_3，V_4 的邻域族 $\mathscr{V}(x)$，再定义开集是可作为它的所有点的邻域的集合，那么同样可得到等价的拓扑空间的定义．用邻域来定义拓扑空间在观念上更自然些，因为"拓扑"这个数学概念是对点与点之间的相邻关系的抽象．但是在逻辑上，用邻域来定义拓扑，不如用开集或闭集来定义拓扑简单，因为 V_4 不是一个显而易见的邻域应该有的性质，而如果没有 V_4，则推不出开集应有的性质．

上面说到的其他的拓扑概念现在也都可用邻域来定义．设 $A\subset X$ 则 x 是 A 的内点，是指 $A\in\mathscr{V}(x)$；A 的内点全体则是 A 的内部 $\operatorname{int}A$．x 是 A 的接触点，是指

$$\forall V\in\mathscr{V}(x),\quad A\cap V\neq\varnothing$$

A 的接触点全体则是 A 的闭包 $\overline{A}$．利用邻域的概念还可把接触点分为两类．设 $x\in\overline{A}$．如果

$$\exists V\in\mathscr{V}(x),\quad V\cap(A\setminus\{x\})=\varnothing,$$

则称 x 是 A 的孤立点；否则，称为 A 的聚点．此外，A 在 B 中稠密也可表达为

$$\forall x\in B,\ \forall V\in\mathscr{V}(x),\ V\cap A\neq\varnothing.$$

一个拓扑空间 X 称为 Hausdorff 空间或分离空间，是指

$$\forall x_1、x_2\in X,\ \exists V_1\in\mathscr{V}(x_1),\ V_2\in\mathscr{V}(x_2),\ V_1\cap V_2=\varnothing. \tag{3.1.1}$$

非分离空间的最简单的例子就是任何平凡拓扑空间．在这种空间中任何点都仅以全空间为其邻域．我们以后几乎只讨论分离拓扑空间．分离性的主要意义在于下面定义的极限的唯一性(习题1)．

一个拓扑空间常常也可只用它的开集族的一部分来定义．设 $\mathscr{B}$ 为拓扑空间 X 的子开集族，即 $\mathscr{B}\subset\mathscr{O}$．如果

$$\forall O\in\mathscr{O},\ \exists B_i\in\mathscr{B}(i\in I),\quad O=\bigcup_{i\in I}B_i, \tag{3.1.2}$$

那么 $\mathscr{B}$ 称为 X 的开集基，也简称为基．类似地，也可定义邻域基

的概念. 设 $\mathscr{U}(x)\subset\mathscr{V}(x)$ 为拓扑空间 X 的点 x 的子邻域族. 如果它满足

$$\forall V\in\mathscr{V}(x),\ \exists U\in\mathscr{U}(x),\ U\subset V, \tag{3.1.3}$$

则 $\mathscr{U}(x)$ 称为 x 的邻域基. 一个拓扑空间的拓扑也可由它的开集基或所有点上的邻域基来决定. 有时利用这点通过已有的拓扑空间来定义新的拓扑空间是比较方便的. 例如, 设 X_i $(i\in I)$ 都是拓扑空间, $\mathscr{O}_i$ $(i\in I)$ 分别是它们的开集族. 如果在 X_i 的乘积集合 $\prod\limits_{i\in I}X_i$ 上定义其开集基为所有形为

$$O_{i_1}\times O_{i_2}\times\cdots\times O_{i_n}\times\prod_{\substack{i\neq i_k\\ k=1,\cdots,n}}X_i,\quad O_{i_k}\in\mathscr{O}_{i_k}(k=1,\ \cdots,\ n) \tag{3.1.4}$$

的集合全体, 这里假定集合的乘积表达形式与其次序无关而仅与下标有关, 那么由此决定的 $\prod\limits_{i\in I}X_i$ 上的拓扑, 称为 X_i $(i\in I)$ 的乘积拓扑, 而赋以这种拓扑的空间 $\prod\limits_{i\in I}X_i$ 则称为 $X_i(i\in I)$ 的拓扑乘积空间. 例如, 作为 $\boldsymbol{R}$ 的乘积集合 $\boldsymbol{R}^n$, 其拓扑就可作为 n 个 $\boldsymbol{R}$ 的乘积拓扑来定义. 但由(3.1.4)决定的 $\boldsymbol{R}^n$ 中的开集都是一些长方块, 而如所周知, $\boldsymbol{R}^n$ 中的通常的开集不必是这种形式的长方块的有限并集.

有了拓扑空间的概念以后, 就可方便地引入极限、连续等概念. 我们先从序列的极限讨论起. 设 $\{x_n\}\subset X$ 为拓扑空间 X 中的序列. 如果存在 $x\in X$, 使得

$$\forall V\in\mathscr{V}(x),\ \exists,\ N\in\boldsymbol{N},\ \forall n\geqslant N,\ x_n\in V,$$

那么称 x 为序列 $\{x_n\}$ 的极限, 或者说 $\{x_n\}$ 收敛于 x, 并记作

$$\lim_{n\to\infty}x_n=x.$$

当 $\{x_n\}\subset A$ 时, 那么

$$x=\lim_{n\to\infty}x_n$$

必定是 $\overline{A}$ 中的元素. 然而, 在一般情形, 如果 $x\in\overline{A}$, 那么不一定存在 $\{x_n\}\subset A$, 使得

$$x=\lim_{n\to\infty}x_n.$$

这是因为 $\mathscr{V}(x)$ 中的元素可能"太多"，使得 x 不足以成为 A 中的任何序列的极限。但当 X 的每一点都有可数邻域基时，就能保证

$$x\in\overline{A}\Leftrightarrow\exists\{x_n\}\subset A,\ \lim_{n\to\infty}x_n=x. \tag{3.1.5}$$

例 1 设 $X=[0,1]$，但 X 上的拓扑通过下列方式定义：空集 $\varnothing$ 以及与 $[0,1]$ 至多相差可数个点的集合称为开集。这样，当 $A=[0,1)$ 时，则任何包含 1 的开集都与 A 相交。故 $\overline{A}=[0,1]$。但是 A 中的任何序列都不可能收敛于 1，因为任何序列都至多包含可数个点，从而由定义，它们都是闭集，只可能收敛于序列中的点。

这个例子虽然是硬造出来的，但却生动地说明了为什么 $\mathscr{V}(x)$ 中的元素"太多"时，x 就不足以成为靠近它的点的序列的极限。因为不管给出怎样的序列，总能找到不包含该序列的邻域。▌

以上例子说明了，希望接触点与极限点之间有(3.1.5)那样的关系，序列极限的概念是不够的。为此，需要定向族或网(net)的概念。集合 D 称为定向集，是指 D 是序集，且满足

$$\forall\alpha、\beta\in D,\ \exists\gamma\in D,\ \alpha\leqslant\gamma,\ \beta\leqslant\gamma.$$

集合 X 上的定向族或网，是指映射 $g: D\to X$，这里 D 是个定向集。仿照序列的记法，我们也把定向族记为 $\{x_\alpha\}_{\alpha\in D}\subset X$。与序列的情形类似，拓扑空间 X 中的定向族 $\{x_\alpha\}_{\alpha\in D}$ 的极限如下定义：如果存在 $x\in X$，使得

$$\forall V\in\mathscr{V}(x),\ \exists\alpha_0\in D,\ \forall\alpha\geqslant\alpha_0,\ x_\alpha\in V,$$

那么称 x 为定向族 $\{x_\alpha\}$ 的极限，或者说 $\{x_\alpha\}$ 收敛于 x，并记作

$$\lim_{\alpha}x_\alpha=x. \tag{3.1.6}$$

有了这个概念以后，就可以说：

$$x\in\overline{A}\Leftrightarrow\exists\text{定向族}\{x_\alpha\}_{\alpha\in D}\subset A,\ \lim_{\alpha}x_\alpha=x. \tag{3.1.7}$$

这里 $\Leftarrow$ 是显然的，为指出 $\Rightarrow$，只需注意到 $\mathscr{V}(x)$ 中的元素全体如果定义 $V_1\leqslant V_2\Leftrightarrow V_1\supset V_2$，则 $\mathscr{V}(x)$ 就成了定向集。于是对于任何 $V\in\mathscr{V}(x)$，取 $x_V\in V\cap A$，则显然有

$$\lim_{V} x_V = x.$$

还可注意到，这里的 V 不需遍经所有 $\mathscr{V}(x)$ 的元素，而只需遍经 x 的邻域基即可．因此当 X 的每点都有可数邻域基时，上述的 x_V 可以只取可数个，即 $\{x_V\}$ 是个序列．这也同时证明了(3.1.5)中所述的结论．

与序列的子列概念相类似，对于定向族，有子定向族的概念．但由于定向集不是全序集，因此，会出现一些序列情形遇不到的怪现象．为了明确起见，我们把定向集记作$(D, \leqslant_D)$．另一个定向集 $(D', \leqslant_{D'})$称为$(D, \leqslant_D)$的子定向集，是指：(i) $D' \subset D$；(ii) $\forall \alpha \in D, \exists \beta \in D', \alpha \leqslant_D \beta$，且 $\beta \leqslant_{D'} \gamma \Rightarrow \alpha \leqslant_{D'} \gamma$．这里最要注意的是 D' 上的序关系 $\leqslant_{D'}$ 可以不必是 D 上原有的序关系．如果 $\leqslant_{D'}$ 就是通过$\leqslant_D$ 导出的，即



$$\forall \alpha、\beta \in D', \alpha \leqslant_{D'} \beta \Leftrightarrow \alpha \leqslant_D \beta,$$

则$(D', \leqslant_{D'})$称为$(D, \leqslant_D)$的*共尾子集*．当$(D', \leqslant_{D'})$是$(D, \leqslant_D)$的子定向集时，定向族 $\{x_\alpha\}_{\alpha \in D'}$ 就称为 $\{x_\alpha\}_{\alpha \in D}$ 的*子定向族*；而当$(D', \leqslant_{D'})$ 是 $(D, \leqslant_D)$ 的共尾子集时，定向族 $\{x_\alpha\}_{\alpha \in D'}$ 就称为 $\{x_\alpha\}_{\alpha \in D}$ 的*共尾子族*．序列的子列仅仅是序列的共尾子族，但是序列还可以有不是子列的子定向族．下列例子说明了，仅有共尾子族(子列)的概念是不够的．

例 2　设

$$X = \left\{\left(\frac{1}{n}, \frac{1}{m}\right)\right\}_{n, m \in N^*} \cup \{(0, 0)\},$$

这里 N^* 表示正整数全体．它的拓扑定义如下：(i) 除 $(0, 0)$ 以外，$\varnothing$和所有单点集是开集；(ii)有下列性质的含$(0, 0)$的集合 V 是开集：除了有限个 n 外，$\left\{\left(\frac{1}{n}, \frac{1}{m}\right)\right\}_{m \in N^*} \backslash V$ 是有限集．容易验证，以 (i)、(ii) 定义的所有开集作为开集族的拓扑空间 X 是 Hausdorff 空间，且$(0, 0)$是$\left\{\left(\frac{1}{n}, \frac{1}{m}\right)\right\}_{n, m \in N^*}$的聚点．但是 X 中的任何不含$(0, 0)$的序列都不收敛于$(0, 0)$．事实上，任何可能收

敛于(0, 0)的序列

$$\{x_k\}\subset\left\{\left(\frac{1}{n},\ \frac{1}{m}\right)\right\}_{n,m\in N^*}$$

不可能只停留在有限个$\left\{\left(\frac{1}{n},\ \frac{1}{m}\right)\right\}_{m\in N^*}$中．而如果在无限个$\left\{\left(\frac{1}{n},\ \frac{1}{m}\right)\right\}_{m\in N^*}$中有$\{x_k\}$的元素，则可以构造一个(0, 0)的不包含无限个$\{x_k\}$中的元素的邻域，以至$\{x_k\}$不可能收敛于(0, 0)．然而，(0, 0)作为$\left\{\left(\frac{1}{n},\ \frac{1}{m}\right)\right\}_{n,m\in N^*}$的聚点，又必定有$\left\{\left(\frac{1}{n},\ \frac{1}{m}\right)\right\}_{n,m\in N^*}$中的元素构成的定向族收敛于(0, 0)．这种定向族一定是序列的$\left\{\left(\frac{1}{n},\ \frac{1}{m}\right)\right\}_{n,m\in N^*}$的子定向族，且它一定不是子列．▌

定向族的概念用来指出某集合是闭集时是很有力的工具．但用来指出某集合是紧集(定义见稍后)却不是很方便的．为此，我们又需要"滤子"的概念，尽管滤子与定向族在某种意义上是两种相互等价的东西．集合X上的子集族$\mathscr{F}$称为X上的滤子(filter)，是指它满足：

i) $\varnothing\notin\mathscr{F}$，且$\mathscr{F}$为非空的子集族；

ii) 如果$F\in\mathscr{F}$，$F\subset G\subset X$，那么$G\in\mathscr{F}$；

iii) 如果F、$G\in\mathscr{F}$，那么$F\cap G\in\mathscr{F}$．

集合X上的子集族$\mathscr{B}$称为X上的滤子基，是指它满足：

i) $\varnothing\notin\mathscr{B}$，且$\mathscr{B}$为非空的子集族；

ii) 如果B_1、$B_2\in\mathscr{B}$，那么存在$B_3\in\mathscr{B}$，使得$B_3\subset B_1\cap B_2$．

显然，每个滤子基$\mathscr{B}$都可唯一地扩张为一个滤子$\mathscr{F}$，使得

$$\forall F\in\mathscr{F},\ \exists B\in\mathscr{F},\ B\subset F.$$

事实上，这时滤子$\mathscr{F}$由所有包含$\mathscr{B}$的元素作为子集的X的子集组成．

滤子与定向族的关系如下：设$\{x_\alpha\}_{\alpha\in D}$为$X$上的定向族，那么令$F_\beta=\{x_\alpha|\alpha\geqslant\beta\}$，则$X$的子集族$\{F_\beta\}_{\beta\in D}$就构成$X$上的一个滤子基．反之，由于$X$上的滤子$\mathscr{F}$关于包含关系就构成一个

定向集，于是对于任何 $F\in\mathscr{F}$，取 $x_F\in F$，$\{x_F\}_{F\in\mathscr{F}}$ 就是 X 上的一个定向族．关于滤子与定向族的各种结论就可通过这样的转换而相互转换．

拓扑空间中的滤子同样也有收敛的概念．设 $\mathscr{F}$ 为拓扑空间 X 中的滤子．如果存在 $x\in X$，使得

$$\mathscr{V}(x)\subset\mathscr{F},$$

那么称 x 为滤子 $\mathscr{F}$ 的极限，或者说 $\mathscr{F}$ 收敛于 x，记作

$$\lim\mathscr{F}=x.$$

对于滤子基可通过它生成的滤子而有同样的概念．它也可如下直接定义：设 $\mathscr{B}$ 为拓扑空间 X 中的滤子基．如果存在 $x\in X$，使得

$$\forall V\in\mathscr{V}(x),\ \exists B\in\mathscr{B},\ B\subset V,$$

那么称 x 为滤子 $\mathscr{B}$ 的极限，或者说，$\mathscr{B}$ 收敛于 x，记作

$$\lim\mathscr{B}=x.$$

与定向族的子定向族相当的概念是比原滤子更细的滤子．设 $\mathscr{F}_1$ 与 $\mathscr{F}_2$ 是集合 X 中的两个滤子．如果 $\mathscr{F}_1\subset\mathscr{F}_2$，那么就称 $\mathscr{F}_2$ 比 $\mathscr{F}_1$ 更细．显然，对于拓扑空间中的两个滤子 $\mathscr{F}_1$ 和 $\mathscr{F}_2$ 如果 $\mathscr{F}_2$ 比 $\mathscr{F}_1$ 更细，那么当 $\mathscr{F}_1$ 收敛时，$\mathscr{F}_2$ 也收敛；反之，则不一定．它们间的关系就像序列及其子列、或更一般的定向族及其子定向族之间的关系．"最细"的滤子，即不存在比它更细的滤子的滤子，称为超滤子．利用 Zorn 引理容易证明，对于集合 X 上的任何滤子，都存在比它更细的超滤子．对于定向族来说，当然也应该有相当的"超定向族"的概念，但较费解，一般不采用．

除了极限外，对于拓扑空间中的滤子还较容易定义聚点的概念．设 $\mathscr{F}$ 为拓扑空间 X 中的滤子．那么

$$\mathrm{cl}(\mathscr{F})=\bigcap_{F\in\mathscr{F}}\overline{F}$$

称为 $\mathscr{F}$ 的聚点集，其中的元素称为 $\mathscr{F}$ 的聚点．对定向族显然也可有类似的概念，但同样不太方便．

现在我们利用滤子的概念来讨论拓扑空间的紧集．分离拓扑空间 X 中的子集 A 称为紧集，是指如果开集族 $O_i\in\mathscr{O}_{\neq}$ $(i\in I)$ 满

是 $A\subset\bigcup_{i\in I}O_i$，那么存在有限个 $i_1, \cdots, i_n\in I$，使得

$$A\subset\bigcup_{k=1}^{n}O_{i_k}.$$

它也可简述为：A 的每一开覆盖具有有限子覆盖．如果 X 本身就是紧集，那么 X 称为紧(拓扑)空间．注意，在这里总要求空间是分离的，以避免“平凡空间也是紧空间”这样的结论．下列定理说明如何用滤子来刻划紧空间．

定理 3.1.1 设 X 是分离拓扑空间．则下列命题是等价的：

i) X 是紧空间；

ii) 设 F_i $(i\in I)$ 为 X 的闭集族．如果任何有限个 F_i 的交是非空的，那么 $\bigcap_{i\in I}F_i\neq\varnothing$；

iii) X 中的任何滤子有聚点；

iv) X 中的任何超滤子是收敛的．

证明 i) $\Leftrightarrow$ ii)．令 $O_i=X\backslash F_i$．则 $X=\bigcup_{i\in I}O_i$ 等价于 $\bigcap_{i\in I}F_i=\varnothing$ 于是 ii) 不过是 i) 的逆否命题．

ii) $\Leftrightarrow$ iii)．ii) $\Rightarrow$ iii) 是显然的．反之，则可把 $\{F_i\}$ 扩充为一个滤子，而导得 $\bigcap_{i\in I}F_i\neq\varnothing$．

iii) $\Leftrightarrow$ iv)．只需注意到如果超滤子 $\mathscr{F}_0$ 有聚点 $a\in X$，那么必定有 $\mathscr{V}(a)\subset\mathscr{F}_0$，即 $\lim\mathscr{F}_0=a$．否则 $\mathscr{V}(a)\cup\mathscr{F}_0$ 将是比 $\mathscr{F}_0$ 更细的滤子．这与 $\mathscr{F}_0$ 是超滤子相矛盾．由此容易看出 iii)、iv) 是等价的．▌

定理 3.1.2(Тихонов) 紧空间的拓扑乘积是紧空间．

证明 设 X_i $(i\in I)$ 都是紧空间．$X=\prod_{i\in I}X_i$ 并赋以乘积拓扑．定义射影映射 P_i：$X\to X_i$ 为

$$\forall x=(x_i)_{i\in I}\in X,\quad P_i(x)=x_i.$$

那么如果 $\mathscr{F}$ 是 $X=\prod_{i\in I}X_i$ 中的超滤子，则必有 $P_i(\mathscr{F})$ 是 X_i 中的超滤子．由定理 3.1.1 的 iv)，存在 $x_i=\lim P_i(\mathscr{F})$．容易验证

$$(x_i)_{i\in I}=\lim\mathscr{F}.\ ▌$$

注 如果用定向族来刻划空间的紧性，那么定理 3.1.1 的

i)~iv)也等价于:

v) X 中的任何定向族有收敛子定向族.

但是紧性不等价于下列两种性质:

(可数紧)任何无限集有聚点;

(列紧)任何序列有收敛子列.

显然,(紧)⇒(可数紧),(列紧)⇒(可数紧). 但(紧)与(列紧)互不包含.

例 3 设 $X=\{f: [0,1]\to[0,1]\}$ 为[0, 1]到[0, 1]的数值函数 f 的全体. 它也可以看作乘积空间

$$X=[0,1]^{[0,1]}=\prod_{t\in[0,1]}[0,1]_t.$$

我们对[0, 1]定义通常的拓扑, 而对 X 赋以乘积拓扑. 则由定理 3.1.2, X 是紧空间. 容易看出, X 中的收敛性就是逐点收敛性. 定义 X 中的点列 $\{f_n\}$ 如下:

$\forall x\in[0,1]$, $f_n(x)=x$ 的二进展开的第 n 位数(0 或 1), $n=1, 2, \cdots$. 则 $\{f_n\}$ 没有收敛子列. 事实上, 如果 $\{f_{n_k}\}\subset\{f_n\}$ 收敛, 即

$$\forall x\in[0,1],\ \lim_{n_k\to\infty}f_{n_k}(x)=f(x).$$

则总能找到某个 $\bar{x}\in[0,1]$, 使得 $\bar{x}$ 的展开的第 n_1 位为 0, 第 n_2 位为 1, 以后总是 0, 1 交叉, 以至 $\lim\limits_{n_k\to\infty}f_{n_k}(\bar{x})$ 不可能存在. 这就说明 X 不是列紧的. 这个例子与例 2 一样, 也是有聚点而无收敛子列的序列的例子.

另一方面, 设 A 是 X 中所有仅在可数个点上不为零的函数的全体. 那么容易看出, A 在 X 中是稠密的, 但 $A\neq X$. 因此, A 不可能是紧集. 然而, A 中的每一序列都可通过对角线方法选出收敛子列, 即 A 是列紧集. ▌

在每一点上都有紧邻域的拓扑空间称为局部紧空间. $\boldsymbol{R}^n$ 对通常的拓扑就是一个局部紧空间. 甚至反之也成立, 即局部紧性是有限维拓扑线性空间的特征(习题三, 13).

有一类特殊的拓扑空间称为距离空间. 集合 X 称为距离空

间是指在其上定义了称为距离的映射 $d: X\times X\to \boldsymbol{R}_+$，满足：

i) $\forall x、y\in X,\quad d(x, y)=0\Leftrightarrow x=y$;

ii) $\forall x、y\in X,\quad d(x, y)=d(y, x)$;

iii) $\forall x、y、z\in X,\quad d(x, y)\leqslant d(x, z)+d(y, z)$.

距离空间 X 中的点 x 的邻域基定义为

$$\mathscr{B}(x)=\left\{B\left(x, \frac{1}{n}\right)\right\}_{n=1,2,\cdots},$$

这里　$B(x, 1/n)=\{y\in X\,|\,d(x, y)<1/n\}$.

因此，距离空间的每个点上都有可数邻域基，以至在距离空间中考虑取极限时，一般只需考虑序列的极限．此外，还可证明，对于距离空间来说，紧、可数紧与列紧是等价的(参看习题三，8)．

最后，两个拓扑空间 X 和 Y 间的映射 $f: X\to Y$ 称为在 $x\in X$ 上连续，是指 $y=f(x)$ 的任何邻域 V 的逆象 $f^{-1}(V)$ 是 x 的邻域．它也可用定向族叙述为：

如果　$\lim\limits_{\alpha} x_\alpha=x$,

那么　$\lim\limits_{\alpha} f(x_\alpha)=f(x)$.

当 X 为距离空间时，定向族还可代替为序列．用滤子来叙述则为：

如果　$\lim \mathscr{F}=x$,

那么　$\lim\limits_{\mathscr{F}} f(x')=y$.

这里 $\lim\limits_{\mathscr{F}} f(x')$ 表示由 Y 中的滤子基 $f(\mathscr{F})$(它一般不是滤子)所生成的滤子的极限．如果 f 在 X 的所有点上都连续，则 f 在 X 上连续．不难证明，$f: X\to Y$ 连续的充要条件为 Y 中的任何开集 O_Y 的逆象 $f^{-1}(O_Y)$ 是 X 中的开集．当开集换为闭集时也成立．由于集合间的映射总是把超滤子变为超滤子，所以由定理 3.1.1 的 iv) 即可得：紧空间的连续象是紧集．如果 $f: X\to Y$ 连续，$f^{-1}: Y\to X$ 存在且连续，那么称 f 为 X 的同胚．

我们经常通过一些拓扑空间与另一个集合间的映射，由要求这些映射的连续性，来定义该集合上的拓扑．例如，设 $A\subset X$ 为拓扑空间 X 的子集，A 上的恒等映射 I 可看作 A 到 X 的映射．

使得 $I: A\to X$ 连续的 A 上的最粗的拓扑称为 X 在 A 上的诱导拓扑. 这种拓扑要求 X 中的开集 O 在 A 中的逆象

$$I^{-1}(O)=O\cap A$$

是 A 中的开集，且由最粗的要求，A 中不再有其他的开集. 在 A 中定义了这样的诱导拓扑后，常称 A 是 X 的拓扑子空间. 又如，设 X_i $(i\in I)$都是拓扑空间. 那么拓扑乘积 $X=\prod_{i\in I}X_i$ 实际上是要求所有射影映射 $P_i: X=\prod_{i\in I}X_i\to X_i$ 都连续的最粗的 X 上的拓扑.

有关拓扑空间的不很完备的补充知识就简述至此.

§ 3.2 拓扑线性空间中的凸集分离定理

现在我们在线性空间中引入拓扑结构，并要求线性结构与拓扑结构相协调，于是就得到了拓扑线性空间的概念. 具体地说，一个(实)线性空间 X 称为拓扑线性空间，是指在 X 中定义了拓扑结构，并满足:

TV_1: $(x, y)\mapsto x+y$ 是 $X\times X$ 到 X 的连续映射;

TV_2: $(\lambda, x)\mapsto\lambda x$ 是 $\mathbf{R}\times X$ 到 X 的连续映射.

这里条件 TV_1、TV_2 就意味着线性结构和拓扑结构的协调.

注 一个线性空间中引入了一种很自然的拓扑结构，并不能保证这种拓扑结构一定与线性结构相协调. 典型的例子就是在线性空间中引入用代数开集作为开集的拓扑; 当空间的维数大于1时，这种拓扑与线性结构不协调. 事实上，对 $\mathbf{R}^2$ 取0的邻域为命题1.3.1的注1中的集合 A. 如果 TV_1、TV_2 满足，那么应该存在 $\varepsilon>0$ 和0的邻域 U、V，使得

$$\lambda U+\mu V\subset A,\quad \forall\lambda、\mu\in(-\varepsilon, \varepsilon).$$

但这是不可能的，因为若取 $x\in U$，$y\in V$，且 x、y 线性无关，那么

$$B=\{z\in\mathbf{R}^2|z=\lambda x+\mu y,\ \lambda,\ \mu\in(-\varepsilon,\ \varepsilon)\}\subset\lambda U+\mu V$$

是以原点为中心的矩形，而 A 中不包括任何这样的矩形.

另一个例子是对线性空间取离散拓扑. ∎

命题 3.2.1 设 X 为拓扑线性空间. 那么对于任何 $x_0\in X$ 和 $\lambda\in\mathbf{R}(\lambda\neq0)$, 映射 $x\mapsto\lambda x+x_0$ 是 X 到 X 的同胚. 特别是, 如果 $\mathscr{V}$ 是 X 的零的邻域族,那么对于任何

$$\lambda\in\mathbf{R}(\lambda\neq0),\quad \lambda\mathscr{V}=\{\lambda V\}_{V\in\mathscr{V}}$$

也是零的邻域族; 对于任何

$$a\in X,\ \mathscr{V}+a=\{V+a\}_{V\in\mathscr{V}}$$

是 a 的邻域族.

证明 事实上,设

$$f(x)=\lambda x+x_0=y,$$

那么

$$f^{-1}(y)=\lambda^{-1}(y-x_0).$$

它们都是 X 到 X 的连续映射, 故 f 是 X 到 X 的同胚. 命题的后半部由连续的定义可得. ∎

命题 3.2.2 设 $\mathscr{U}$ 为拓扑线性空间 X 的零邻域基. 那么对于任何 $U\in\mathscr{U}$, 有

i) $0\in U^i$, 即 U 是吸收集;

ii) 存在 $V\in\mathscr{U}$, 使得 $V+V\subset U$;

iii) 存在对称的零邻域 W, 即 $W=-W$, 使得 $W\subset U$. 反之, 如果 X 的滤子基 $\mathscr{U}$ 满足 i)~iii) (iii) 更改为存在对称的 $W\in\mathscr{U}$), 那么 X 上可定义以 $\mathscr{U}$ 为零邻域基的与线性结构相协调的拓扑.

证明 i) 对于任何 $h\in X$, $H(\lambda)=\lambda h$ 是 $\mathbf{R}$ 到 X 的连续映射, 从而对于任何 $U\in\mathscr{U}$, 存在 $\lambda=0$ 的邻域 $(-\varepsilon,\ \varepsilon)$, 使得

$$H((-\varepsilon,\ \varepsilon))\subset U,$$

即存在 $\varepsilon>0$, 使得 $(-\varepsilon h,\ \varepsilon h)\subset U$. 因此 $0\in U^i$.

ii) 因为 $g(x,\ y)=x+y$ 是 $X\times X$ 到 X 的连续映射, 故对于任何 $U\in\mathscr{U}$, 存在 X 的零邻域 V_1、V_2, 满足 $V_1+V_2\subset U$. 取 $V\in\mathscr{U}$, 使得 $V\subset V_1\cap V_2$, 则 $V+V\subset U$.

iii) 因为 $G(\lambda,\ x)=\lambda x$ 是 $\mathbf{R}\times X$ 到 X 的连续映射, 故对于任何 $U\in\mathscr{U}$, 存在 $\lambda=0$ 的邻域 $(-\varepsilon,\ \varepsilon)$ 和 X 的零邻域 V, 使得

$$\forall\lambda\in(-\varepsilon,\ \varepsilon),\quad \lambda V\subset U.$$

令
$$W=\frac{\varepsilon}{2}V\cap\left(-\frac{\varepsilon}{2}V\right)$$
则 $W=-W$，且 $W\subset U$.

反之，我们在 X 上定义以 $\mathscr{U}$ 为零邻域基的拓扑(对于任何 $a\in X$，以 $\mathscr{U}+a$ 作为 a 的邻域基)，那么首先易证 $(x, y)\mapsto x+y$，$(\lambda, x)\mapsto\lambda x$ 在 $x=y=0$、$\lambda=0$ 时连续；然后再利用平移可得到它们处处连续. ∎

注 令 $\mathscr{U}$ 为线性空间 X 中所有包含 0 的凸代数开集. 那么 $\mathscr{U}$ 满足 i)、ii)、iii). 于是 X 上可定义以这样的 $\mathscr{U}$ 为零邻域基的与线性结构相协调的拓扑. 这时前一注中的问题不再出现，因为集合 A 不再是零邻域. ∎

推论 任何拓扑线性空间有闭的对称的零邻域基.

证明 事实上，如果 $x\in\overline{V}$，则
$$(x-V)\cap V\neq\varnothing,$$
即
$$x\in V+V,$$
从而 $\overline{V}\subset V+V$. 这样由 ii)、iii)，对每一 $U\in\mathscr{U}$，存在一闭的对称的邻域 $W\subset U$. ∎

命题3.2.3 设 $\mathscr{U}$ 为拓扑线性空间 X 的零邻域基. 那么 X 是分离空间，当且仅当
$$\bigcap_{U\in\mathscr{U}}U=\{0\}. \tag{3.2.1}$$

证明 如果 X 是分离的，那么
$$\forall x\in X,\ x\neq 0,\ \exists U\in\mathscr{U},\ x\notin U,$$
从而(3.2.1)成立.

反之，如果(3.2.1)成立，则对于任何 x、$y\in X$，$x\neq y$，存在 $U\in\mathscr{U}$，使得 $x-y\notin U$. 由命题 3.2.2 的 ii)、iii)，存在对称的零邻域 V，使得 $V+V\subset U$. 于是必须有
$$(x+V)\cap(y+V)=\varnothing,$$
否则，将存在 $z\in(x+V)\cap(y+V)$，使得
$$x-y=(z-y)-(z-x)\in V-V=V+V\subset U,$$
导致矛盾. ∎

推论　拓扑线性空间是分离的当且仅当$\{0\}$是闭集(从而每一单点集也是闭集). ■

"当"的部分是命题 3.2.2 的推论与命题 3.2.3 合在一起的结果;"仅当"部分可由定义直接证明(习题三,2).

命题 3.2.4　设 A 为拓扑线性空间 X 中的线性子空间(仿射集、锥、凸集). 那么 A 在 X 中的闭包 $\overline{A}$ 也是线性子空间(仿射集锥、凸集).

证明　以凸集为例,其他情况是类似的. 设 x、$y\in\overline{A}$, 则存在定向族 $\{x_\alpha\}$、$\{y_\beta\}\subset A$,使得 $x_\alpha\to x$, $y_\beta\to y$. 对于任何 $\lambda\in[0,1]$, $\{(1-\lambda)x_\alpha+\lambda y_\beta\}\subset A$ 也是定向族, 其定向指标集为 $\{(\alpha,\ \beta)\}$, 且

$$(\alpha,\ \beta)\leqslant(\alpha',\ \beta')$$

定义为 $\alpha\leqslant\alpha'$、$\beta\leqslant\beta'$ 同时成立. 显然有

$$(1-\lambda)x_\alpha+\lambda y_\beta\to(1-\lambda)x+\lambda y,$$

故

$$(1-\lambda)x+\lambda y\in\overline{A}.$$

由 λ、x、y 的任意性, $\overline{A}$ 也是凸集. ■

下面讨论拓扑线性空间上的线性形式. 由于空间有了拓扑, 我们可以谈论线性形式的连续性. 所谓线性形式 $x^*\in X^*$ 是连续的, 是指 X 到 $\mathbf{R}$ 的映射 $x\mapsto\langle x^*,\ x\rangle$ 是连续的.

命题 3.2.5　设 x^* 为拓扑线性空间 X 上的线性形式. 那么 x^* 连续的充要条件为

$$V_1=\{x\in X|\ \ |\langle x^*,\ x\rangle|<1\}\tag{3.2.2}$$

是 X 的零邻域.

证明　x^* 在 $x=0$ 上连续的充要条件为对于任何 $\varepsilon>0$,

$$V_\varepsilon=\{x\in X|\ \ |\langle x^*,\ x\rangle|<\varepsilon\}$$

是零邻域. 但 V_ε 是零邻域等价于 $V_1=V_\varepsilon/\varepsilon$ 是零邻域,故命题的条件等价于 x^* 在 $x=0$ 上连续. 由 x^* 的线性性, 易证 x^* 在 $x=0$ 上连续等价于在 X 上连续. ■

注　对于任何 $x^*\in X^*$, 形为(3.2.2)的集合 V_1 显然是代数开凸集. 因此,如果定义所有含 0 的代数开凸集为零邻域基, 那么所有线性形式都是连续的. ■

定理 3.2.6 设 x^* 为拓扑线性空间 X 上的线性形式. 那么 x^* 连续的充要条件为超平面

$$H_0=\{x\in X \mid \langle x^*, x\rangle=0\}$$

是闭的. 如果 H_0 不是闭的, 那么 H_0 必定在 X 中是稠密的, 即

$$\overline{H}_0=X.$$

证明 如果 x^* 连续, 那么 H_0 作为 $\mathbf{R}$ 中的闭集 $\{0\}$ 的逆象也是闭的. 反之, 如果 H_0 是闭的, 那么对于任何 $x_1\notin H_0$, 存在零邻域 U, 使得 $(x_1+U)\cap H_0=\varnothing$. 不妨设 $\langle x^*, x_1\rangle=1$, 且 U 对称. 那么必须有

$$U\subset V_1=\{x \mid |\langle x^*, x\rangle|<1\}.$$

否则, 将存在 $\tilde{x}\in U$, 使得

$$|\langle x^*, \tilde{x}\rangle|\geqslant 1.$$

由 U 对称, 可假设

$$\langle x^*, \tilde{x}\rangle\leqslant -1.$$

但又由 $0\in U'$, 可得 $[0, \tilde{x}]\subset U$, 从而存在 $\tilde{x}_1\in[0, \tilde{x}]$, 使得

$$\langle x^*, \tilde{x}_1\rangle=-1,$$

于是

$$\langle x^*, x_1+\tilde{x}_1\rangle=0,$$

与 $(x_1+U)\cap H_0=\varnothing$ 矛盾. 这样一来, $U\subset V_1$, 因此 V_1 也是零邻域. 由命题 3.2.5, x^* 连续.

如果 H_0 不是闭集, 那么 $\overline{H}_0$ 是包含 H_0 为真子空间的 X 的子空间(命题 3.2.4). 令 $x_0\in\overline{H}_0\backslash H_0$, 则 $\langle x^*, x_0\rangle\neq 0$, 从而对于任何 $x\in X$, 存在 $\lambda\in\mathbf{R}$, 使得

$$\langle x^*, x\rangle=\lambda\langle x^*, x_0\rangle=\langle x^*, \lambda x_0\rangle,$$

即 $x-\lambda x_0\in H_0$, 或 $x\in H_0+\lambda x_0\subset\overline{H}_0$. 由 x 的任意性, 这说明

$$\overline{H}_0=X. \quad \blacksquare$$

拓扑线性空间 X 的连续线性形式的全体称为 X 的拓扑对偶. 由于以后很少同时使用代数对偶和拓扑对偶, 我们仍记 X 的拓扑对偶为 X^*, 其中的元素为 x^*. 仅在有需要区别时再另作说明. 即使在以后仍把 X^* 理解为代数对偶, 也不会产生大问题, 因为我们已经看到, 对于 X 的特殊的拓扑, 即以代数开凸集作为开集基, 拓

扑对偶与代数对偶是可以恒同的.

有了上述准备后，我们来指出拓扑线性空间中的凸集分离定理. 与以前不同的是，现在要求分离超平面是闭的，即所对应的线性形式是连续的. 为此，先讨论凸集的代数内部与拓扑内部的关系.

命题 3.2.7 设K为拓扑线性空间 X 中的凸集，且K的拓扑内部 $\text{int}\, K\neq\varnothing$. 那么

i) $\text{int}\, K=K^i$;

ii) $\overline{K}=K^c$.

证明 i) 设 $x\in\text{int}\, K$，则存在零邻域 U，使得

$$x+U\subset K.$$

由命题 3.2.2，$0\in U^i$，从而

$$x\in(x+U)^i\subset K^i.$$

因此，$\text{int}\, K\subset K^i$. 反之，设 $\bar{x}\in K^i$，$x_0\in\text{int}\, K\subset K^i$. 则由定理 1.3.6，存在 $x_1\in K$，使得 $\bar{x}\in[x_0, x_1)$. 而

$$[x_0, x_1)\subset\text{int}\, K.$$

事实上，由 $x_0\in\text{int}\, K$，存在零邻域 U，使得 $x_0+U\subset K$. 由 K 是凸集，故有

$$\forall\lambda\in[0, 1),$$

$$(1-\lambda)(x_0+U)+\lambda x_1=(1-\lambda)x_0+\lambda x_1+(1-\lambda)U\subset K.$$

因为$(1-\lambda)U$ 还是零邻域，故

$$[x_0, x_1)=\{x\,|\,x=(1-\lambda)x_0+\lambda x_1,\ \lambda\in[0, 1)\}\subset\text{int}\, K.$$

特别是 $\bar{x}\in\text{int}\, K$，从而 $K^i\subset\text{int}\, K$ 以至 $K^i=\text{int}\, K$.

ii) 显然有 $K^c\subset\overline{K}$. 反之，如果 $x_1\in\overline{K}$，$x_0\in\text{int}\, K$，那么与上面一样可证明 $[x_0, x)\subset\text{int}\, K$，从而 $x_1\in K^c$. ▌

由命题 3.2.7 和定理 1.4.6、定理 1.4.8，即可得以下定理.

定理 3.2.8 (凸集分离定理)设 A、B 为拓扑线性空间 X 中的两个非空集合，$A-B$ 为凸集，$\text{int}(A-B)\neq\varnothing$，且

$$0\notin\text{int}(A-B).$$

那么 A、B 可用闭超平面分离，0 与 $\text{int}(A-B)$可用闭超平面严格

分离．如果此外还有 $0\notin\overline{(A-B)}$，那么 A、B 可用闭超平面强分离．∎

这里只需注意到把 0 与 $\mathrm{int}(A-B)$ 严格分离的超平面不包含非空开集 $\mathrm{int}(A-B)$ 的点，因而它不可能在 X 中稠密，所以只能是闭的．其他都可由以前的结果得到．

推论 1 设 A、B 为拓扑线性空间 X 中的两个非空凸集．

$$\mathrm{int}\,A\neq\varnothing,\quad \mathrm{int}\,A\cap B=\varnothing,$$

那么 A、B 可用闭超平面分离，$\mathrm{int}\,A$ 与 B 可用闭超平面严格分离．∎

与以前一样，这时可证明

$$\mathrm{int}(A-B)=\mathrm{int}\,A-B.$$

推论 2 设 A 为拓扑线性空间 X 中的凸集，且

$$\mathrm{int}\,A\neq\varnothing.$$

那么对于任何 $x\notin\mathrm{int}\,A$，存在连续的 $x^*\in X^*$，使得

$$\langle x^*,\ x\rangle<\langle x^*,\ y\rangle,\ \forall y\in\mathrm{int}\,A.\ ∎$$

推论 3 设 K 为拓扑线性空间 X 中的闭凸集，且

$$\mathrm{int}\,K\neq\varnothing.\quad \sigma_K(x^*)=\sup_{x\in K}\langle x^*,\ x\rangle$$

为其承托函数，这里 $x^*\in X^*$，X^* 为 X 的拓扑对偶．那么

$$K=\{x\in X\,|\,\langle x^*,\ x\rangle\leqslant\sigma_K(x^*),\ \forall x^*\in X^*\}.\ ∎$$

推论 4 设 K 为分离拓扑线性空间 X 中的闭凸集，且

$$\mathrm{int}\,K\neq\varnothing.$$

C 为 X 中的紧集，$K\cap C=\varnothing$．那么存在闭超平面使 K、C 强分离．

证明 可以指出（留给读者作为练习，参看习题三．9.ii），这时 $K-C$ 为有非空内部的闭凸集，且

$$0\notin K-C.\ ∎$$

与以前相比较，是否可提出"相对拓扑内部"的概念，把所有结论的条件都减弱为"相对拓扑内部"？确实可以提出相对拓扑内部的概念如下：设 A 为拓扑线性空间 X 中的子集．A 的相对拓扑

内部是指：

$$\mathrm{ri}\,A=\{x\in X \mid \exists \text{ 零邻域 } U,\ x+(U\cap\overline{\mathrm{aff}\,A})\subset A\}.$$

对于这个概念，不难指出命题 3.2.7 确实也能改进为如下命题。

命题 3.2.7′　设 K 为拓扑线性空间 X 中的凸集，且

$$\mathrm{ri}\,K\neq\varnothing.$$

那么

i) $\mathrm{ri}\,K=K^{ri}$;

ii) $\overline{K}=K^{c}$. ▌

但是定理 3.2.8 及其推论却无法一样改进．问题在于无法保证任意一个闭仿射集一定能扩充为一个闭超平面，因为存在某些拓扑线性空间，其上不存在任何闭超平面，即不存在任何非零连续线性形式．事实上，由定理 3.2.8，可以得到以下命题：

命题 3.2.9　拓扑线性空间 X 中存在闭超平面或非零连续线性形式的充分必要条件为：X 中存在凸开真子集．

证明　如果 x^* 是 X 上的非零连续线性形式，那么开半空间

$$H_+=\{x\in X \mid \langle x^*,\ x\rangle<0\}$$

是 X 中的凸开真子集．

反之，如果 A 是 X 中凸开真子集，那么存在 $x\in X\backslash A$．由定理 3.2.8，存在闭超平面，使 x 与 A 分离．▌

然而，不包含任何凸开真子集的拓扑线性空间是存在的．见下面的例子：

例　设 $S(0,\ 1)$ 为 $[0,\ 1]$ 区间上所有分段连续函数的全体．则 $S(0,\ 1)$ 显然为线性空间．定义

$$\|x\|=\int_0^1\frac{|x(t)|}{1+|x(t)|}\,dt,$$

并定义 $S(0,\ 1)$ 中的零邻域基为

$$\mathscr{U}=\left\{U_n \;\middle|\; U_n=\left\{x\in S(0,\ 1) \;\middle|\; \|x\|<\frac{1}{n}\right\},\ n=1,\ 2,\ \cdots\right\}.$$

则 $S(0,\ 1)$ 就成为拓扑线性空间．在此指出 $S(0,\ 1)$ 中没有任何凸开真子集．事实上，设 A 为 $S(0,\ 1)$ 中的凸开集，不妨设 $0\in A$．那么必须存在某个 n，使得 $U_n\subset A$．任取 $x\in S(0,\ 1)$，定

又

$$x_k(t)=\begin{cases} nx(t), & \text{当 } t\in\left[\dfrac{k-1}{n},\ \dfrac{k}{n}\right]; \\ 0, & \text{当 } t\notin\left[\dfrac{k-1}{n},\ \dfrac{k}{n}\right]. \end{cases}$$

那么

$$\|x_k\|=\int_0^1\frac{|x_k(t)|}{1+|x_k(t)|}dt=\int_{\frac{k-1}{n}}^{\frac{k}{n}}\frac{n|x(t)|}{1+n|x(t)|}dt\leqslant\frac{1}{n},$$

从而 $x_k\in A$. 但 A 又是凸集，故又有

$$x=\frac{1}{n}(x_1+\cdots+x_n)\in A.$$

由 x 的任意性，即得

$$A=S(0,\ 1).\ \blacksquare$$

命题 3.2.9 和上述例子说明了定理 3.2.8 及其推论无法象以前那样改进，因为甚至分离拓扑线性空间中任意两个不同点都不一定能用闭超平面分离. 下列定理 1.4.10 的变形再次说明了对于凸集分离定理来说，凸开真子集存在的必要性.

定理 3.2.10 拓扑线性空间 X 中的两个非空凸集 A、B 可用闭超平面强分离的充要条件为：存在 X 中的凸零邻域 V，使得

$$(A+V)\cap B=\varnothing.\ \blacksquare$$

下节中要讨论具有凸零邻域基的拓扑线性空间. 可以看出，只有这种空间才能使拓扑对偶、凸集分离定理等比较有意义.

§3.3 局部凸空间

具有凸零邻域基的拓扑线性空间称为局部凸空间. 对于局部凸空间来说，首先有以下一些有意义的结论：

命题 3.3.1 设 X 为有非平凡拓扑的局部凸空间. 那么 X 中有闭超平面，即有非零连续线性形式. $\blacksquare$

命题 3.3.2 设 X 为分离局部凸空间. 那么对于任何 x_1、$x_2\in X$，存在连续线性形式 x^*，使得

$$\langle x^*, x_1\rangle \neq \langle x^*, x_2\rangle. \ ▮$$

因为这时存在凸零邻域 U，使

$$(x_1+U)\cap(x_2+U)=\varnothing.$$

命题 3.3.3　设 X 为局部凸空间．$A\subset X$ 为闭仿射真子集，那么存在包含 A 的闭超平面．

证明　事实上，这时存在 $x_0\in X$，使得 $x_0\notin A$．由于 A 是闭集，故存在凸零邻域 U，使得

$$(x_0+U)\cap A=\varnothing.$$

由定理 3.2.8 的推论 2 或定理 3.2.10，存在连续线性形式，使得

$$\langle x^*, x\rangle<\langle x^*, x_0\rangle, \quad \forall x\in A.$$

但由 A 是仿射集，易证 x^* 在 A 上是常数，即 A 在某个由 x^* 决定的闭超平面上．▮

有了命题 3.3.3 以后，如果假定 X 为局部凸空间，那么定理 3.2.8 及其推论中的拓扑内部都可改为相对拓扑内部．尤其是可得到定理 1.4.8 的推广如下：

定理 3.3.4　设 A、B 为局部凸空间 X 中的凸集，$\mathrm{ri}\,A\neq\varnothing$，$\mathrm{ri}\,B\neq\varnothing$，且 $\mathrm{ri}\,A\cap\mathrm{ri}\,B=\varnothing$．那么，$A$、$B$ 可用闭超平面分离，$\mathrm{ri}\,A$、$\mathrm{ri}\,B$ 可用闭超平面严格分离．▮

我们说定理 3.3.4 为定理 1.4.8 的推广，是因为定理 1.4.8 可以看作对线性空间 X 赋以"最细的局部凸拓扑"后的定理 3.3.4．所谓"最细的局部凸拓扑"是指以所有吸收凸集作为零邻域基的拓扑．由命题 3.2.7 可知，X 上的任何局部凸拓扑都比它粗．

下列结果说明，有了局部凸拓扑，有时甚至连相对拓扑内部都不用考虑．

定理 3.3.5　设 K 为分离局部凸空间 X 中的闭凸集，C 为 X 中的紧凸集，$K\cap C=\varnothing$．那么存在闭超平面使 K、C 强分离．▮

这是因为 $K-C$ 是闭凸集，而 $0\notin K-C$，从而存在凸零邻域 U，使得 $U\cap(K-C)=\varnothing$．于是此定理由定理 3.2.10 而得到．

推论　设 K 为分离局部凸空间 X 中的集合，

$$\sigma_K(x^*)=\sup_{x\in K}\langle x^*, x\rangle$$

为其承托函数(定义在 X 的拓扑对偶 X^* 上)．那么 K 的闭凸包(含 K 的最小闭凸集)为

$$\overline{co}\,K=\{x\in X\,|\,\langle x^*, x\rangle\leqslant\sigma_K(x^*),\ \forall x^*\in X^*\}.$$

证明　右端是包含 K 的闭凸集，故也包含左端．反之，如果

$$x_0\notin\overline{co}\,K,$$

则由定理 3.3.5，存在 $x_0^*\in X^*$，使得

$$\langle x_0^*, x_0\rangle>\sup_{x\in\overline{co}K}\langle x_0^*, x\rangle\geqslant\sigma_K(x_0^*),$$

即 x_0 也不是右端的元素．∎

现在指出在 § 2.3 中提到过的事实：局部凸空间可用一族半范数来刻划．我们记得，线性空间 X 上的函数 $p: X\to\mathbf{R}_+$ 称为 X 上的半范数，是指它是满足 $p(x)=p(-x)$ 的次线性函数，它也可直接定义为满足下列三条件的函数：

i) $\forall x\in X,\ p(x)\geqslant 0$;

ii) $\forall x\in X,\ \forall\lambda\in\mathbf{R},\ p(\lambda x)=|\lambda|p(x)$;

iii) $\forall x、y\in X,\ p(x+y)\leqslant p(x)+p(y)$.

由命题 2.3.4、2.3.5 和 p 的对称性可知，每个 X 上的半范数 p 总对应着 X 中的一个对称吸收凸集 U，它们间的关系如下：

$$U^i=\{x\in X\,|\,p(x)<1\};$$

$$U^c=\{x\in X\,|\,p(x)\leqslant 1\};$$

$$p(x)=\inf\{\alpha>0\,|\,[0, x/\alpha]\subset U\}.$$

且任何界于上述 U^i 和 U^c 之间的集合也都满足这些关系．而由命题 3.2.2 和定义，局部凸空间无非就是用一族对称吸收凸集来确定其拓扑结构的空间．因此，它也一定能用一族半范数来确定其拓扑．更确切地说，有以下定理．

定理 3.3.6　设 $\{p_i\}_{i\in I}$ 为线性空间 X 上的一族半范数．则 X 上存在使每个 p_i 都连续的与线性结构相协调的最粗的拓扑结构．对于这一拓扑，X 为局部凸空间，且以下形式的集合构成了 X 的闭零邻域基：

$$\{x\in X \mid \max_{1\leqslant k\leqslant n} p_{i_k}(x)\leqslant\varepsilon\},\tag{3.3.1}$$

$$\varepsilon>0,\ n\in\mathbf{N},\ i_k\in I\ (k=1,2\cdots,n).$$

证明 如果对于 X 上的某一拓扑，所有 p_i 连续，那么

$$U_{i\varepsilon}=p_i^{-1}([-\varepsilon,\varepsilon])=\{x\in X \mid p_i(x)\leqslant\varepsilon\},\ \varepsilon>0,\ i\in I,$$

必须是 X 中的闭零邻域．从而

$$\{x\in X \mid \max_{1\leqslant k\leqslant n} p_{i_k}(x)\leqslant\varepsilon\}=\bigcap_{1\leqslant k\leqslant n}\{x\in X \mid p_{i_k}(x)\leqslant\varepsilon\}=\bigcap_{1\leqslant k\leqslant n}U_{i_k\varepsilon},$$

也必须是 X 中的闭零邻域．由 p_i 都是半范数，故所有的 $U_{i\varepsilon}$ 都是对称吸收凸集；特别是，由形为(3.3.1)的集合全体构成满足命题 3.2.2 条件的滤子基(其中 ii)由凸性取 $V=U/2$ 可得)．因此，根据命题 3.2.2，它们可用来作为(闭)零邻域基定义 X 上的与线性结构相协调的拓扑．这样定义的拓扑容易验证是使所有 p_i 都连续的最粗的拓扑．▌

由定理 3.3.6 所定义的 X 上的拓扑，称为 X 的由半范数 $\{p_i\}_{i\in I}$ 所确定的局部凸拓扑．而由对称吸收凸集与半范数的相互对应，任何局部凸拓扑也必定可由这种方式来定义．

下列命题是命题 3.2.3 的推论：

命题 3.3.7 设 X 为局部凸空间，其拓扑由半范数 $\{p_i\}_{i\in I}$ 所确定．那么 X 是分离空间，当且仅当

$$\forall x\in X,\ x\neq 0,\ \exists i\in I,\ p_i(x)>0.\tag{3.3.2}$$ ▌

证明只需注意到条件(3.3.2)等价于由(3.3.1)所决定的 X 的零邻域基的交为$\{0\}$即可．

命题 3.3.7 显示了用半范数族来表达空间的一般性质的便利．下面还可举出一些类似的例子．例如：

X 中的定向族$\{x_\alpha\}$满足

$$\lim_\alpha x_\alpha=x\Leftrightarrow\forall i\in I,\ \lim_\alpha p_i(x_\alpha-x)=0.$$

X 上的线性形式 x^* 连续 $\Leftrightarrow\exists i_1,\cdots,i_n\in I,\ \forall x\in X,\ \langle x^*,x\rangle\leqslant\max_{1\leqslant k\leqslant n} p_{i_k}(x)$.

这里 $\Leftarrow$ 部分不太明显，其证明留给读者作为练习(利用命题

3.2.5, 也可参看定理 3.5.1).

下面是一些新的概念,在局部凸情形,它们用半范数族来表述是特别方便的.

设 $\{x_\alpha\}_{\alpha\in D}$ 为拓扑线性空间 X 中的定向族. 如果对于 X 的任何零邻域 U, 存在 $\alpha_0\in D$, 使得

$$\forall\alpha,\ \beta\geqslant\alpha_0,\ x_\alpha-x_\beta\in U,$$

那么 $\{x_\alpha\}$ 称为 X 的 Cauchy 定向族. 类似地也可定义 Cauchy 滤子. 对于有半范数族 $\{p_\iota\}_{\iota\in I}$ 的局部凸空间 X 来说,

$\{x_\alpha\}_{\alpha\in D}$ 是 Cauchy 定向族 $\Leftrightarrow\forall\iota\in I$, $\{p_\iota(x_\alpha)\}_{\alpha\in D}$ 是 $\mathbf{R}$ 中的 Cauchy 定向族.

拓扑线性空间 X 称为完备的,是指 X 中的每一 Cauchy 定向族都是收敛的. 对于局部凸空间来说:

$$X \text{ 完备} \Leftrightarrow \forall \text{Cauchy 定向族 } \{x_\alpha\}_{\alpha\in D},\ \exists x\in X,\ \forall\iota\in I$$

$$\lim_\alpha p_\iota(x_\alpha-x)=0.$$

拓扑线性空间 X 中的集合 A 称为有界集, 是指对于 X 的任何零邻域 U, 存在 $\lambda>0$, 使得 $\lambda A\subset U$. 对于局部凸空间来说:

$$A \text{ 有界} \Leftrightarrow \forall\iota\in I,\ \{p_\iota(x)\}_{x\in I} \text{ 有界}.$$

最常用的局部凸空间是赋范空间和赋距空间. 赋范空间即由一个范数(满足$[x=0\Leftrightarrow p(x)=0]$的半范数 p, 通常记作 $\|\cdot\|$)来确定拓扑的局部凸空间;(局部凸)赋距空间则是可由距离来确定拓扑的局部凸空间. 由于范数也是距离,故赋范空间一定是赋距空间,但反之不一定. 完备的赋范空间称为 Banach 空间, 完备的(局部凸)赋距空间称为 Fréchet 空间. 下面指出,局部凸空间为赋范空间和赋距空间的充要条件.

定理 3.3.8(Колмогоров) 分离局部凸空间可赋范的充要条件为存在有界零邻域.

证明 设 X 为赋范空间, $p(x)=\|x\|$为 X 上的范数. 那么X 中的单位闭球

$$\overline{B}=\{x\in X\mid\|x\|\leqslant1\}$$

就是 X 的有界零邻域.

反之，设 X 为分离局部凸空间，U 为其有界零邻域. 不妨设 U 为对称闭凸集，从而

$$p_U(x)=\inf\{\alpha>0 \mid x\in\alpha U\}$$

为半范数. 由 U 有界，故对于 X 的任何零邻域 V，存在 $\lambda>0$，使得 $\lambda U\subset V$. 因此，

$$\{\varepsilon U\}_{\varepsilon>0}=\{x\in X \mid p_U(x)\leqslant\varepsilon\}_{\varepsilon>0}$$

构成 X 的零邻域基. 另一方面，由 X 分离，故根据命题 3.2.3，有

$$\{x\in X \mid p_U(x)=0\}=\bigcap_{\varepsilon>0}\varepsilon U=\{0\},$$

即 $p_U(x)$ 是 X 中的范数. 这样，X 的拓扑由范数 p_U 所确定. ∎

注 定理 3.3.8 显然也可叙述为：

定理 3.3.8′ 分离拓扑线性空间可赋范的充要条件为存在有界凸零邻域. ∎

定理 3.3.9 分离局部凸空间可赋距的充要条件为其拓扑可由可数半范数族确定.

证明 设 X 为局部凸赋距空间，d 为 X 上的距离. 那么

$$V_n=\left\{x\in X \,\middle|\, d(0,\ x)\leqslant\frac{1}{n}\right\}\quad(n=1,\ 2,\ \cdots)$$

构成 X 上的零邻域基. 由 X 的局部凸性，故对于任何 V_n，存在对称凸零邻域 U_n，满足 $U_n\subset V_n$. 对应每个 U_n 的半范数 p_n 全体就确定了 X 的局部凸拓扑.

反之，如果分离局部凸空间 X 的拓扑由可数个半范数 p_n $(n=1,\ 2,\ \cdots)$所确定，则由命题 3.3.7，有

$$x=0\Leftrightarrow\forall n,\ p_n(x)=0.$$

从而易证

$$d(x,\ y)=\sum_{n=1}^{\infty}2^{-n}p_n(x-y)/[1+p_n(x-y)],\ \forall x,\ y\in X,$$

是 X 上的距离，且对平移不变，即

$$\forall z\in X,\ d(x+z,\ y+z)=d(x,\ y)。$$

不难看出，对于任何 $\{x_\alpha\}\subset X$，有

$$d(x_\alpha,\ 0)\to0\Leftrightarrow\forall n,\ p_n(x_\alpha)\to0.$$

这说明了这个距离 d 所定义的拓扑与原来的是一致的. ▌

注 一个距离线性空间，即可用距离确定拓扑的拓扑线性空间,不一定是局部凸的. 比定理 3.3.9 更一般的结果为：分离拓扑线性空间是距离线性空间的充要条件为存在可数零邻域基. 它的证明较困难. 参看例如，G. Köthe, Topologische Lineare Räume, 1, Springer-Verlag, Berlin, 1960, 166 页. ▌

在这节的最后还应简要指出，任何分离局部凸空间都可用赋范空间(甚至 Banach 空间)来表示.

设 X 为由半范数族 $\{p_i\}_{i\in I}$ 确定拓扑的局部凸空间. 在 X 中定义下列等价关系"$\overset{i}{\sim}$":

$$x \overset{i}{\sim} y \Leftrightarrow p_i(x-y)=0.$$

那么由于 p_i 的零空间

$$N_i=\{x\in X\,|\,p_i(x)=0\}$$

是线性空间,故商空间

$$X_i=X/\overset{i}{\sim}=X/N_i$$

也是线性空间，且由 p_i 在 X 上确定的拓扑所得到的 X_i 上的商拓扑,即使 $X\to X_i$ 的典型映射连续的最细的拓扑,必定使 X_i 成为赋范空间. 对这样的赋范空间 X_i, 作拓扑乘积空间

$$\tilde{X}=\prod_{i\in I}X_i,$$

并令 $g_i: X\to X_i=X/N_i$ 为典型映射,那么映射

$$g(x)=\prod_{i\in I}g_i(x)\in\tilde{X}$$

是 X 到 $\tilde{X}$ 的线性单射,即 $g: X\to\tilde{X}$ 是线性映射,且

$$x=0\Leftrightarrow g(x)=0,$$

后者是因为 X 分离,故

$$x=0\Leftrightarrow \forall i\in I,\ p_i(x)=0.$$

这样在同构意义下, X 可看作 $\tilde{X}$ 的子空间. 另一方面，由 X 中的收敛性的定义:

$$x_\alpha\to x\Leftrightarrow \forall i\in I,\quad p_i(x_\alpha-x)\to 0,$$

以及由乘积拓扑的定义，可知 X 上的拓扑可以看作 $\tilde{X}$ 上的拓扑的诱导拓扑，从而就得到以下定理。

定理 3.3.10 任何分离局部凸空间总拓扑同构于赋范空间的拓扑乘积的子空间. ▮

注 由于任何赋范空间总可看成它的完备化 Banach 空间的子空间，上述定理中的"赋范空间"还可改为"Banach 空间. ▮

§3.4 对偶系和极化拓扑

在上节中已指出，如果 X 是分离局部凸空间，那么它的拓扑对偶 X^* 是"非平凡的"，即 $\neq\{0\}$，并且还满足：

$$\forall x_1、x_2\in X,\ x_1\neq x_2,\ \exists x^*\in X^*,\ \langle x^*, x_1\rangle\neq\langle x^*, x_2\rangle. \quad (3.4.1)$$

同时，下列关系由定义而得到：

$$\forall x_1^*、x_2^*\in X^*,\ x_1^*\neq x_2^*,\ \exists x\in X,\ \langle x_1^*, x\rangle\neq\langle x_2^*, x\rangle. \quad (3.4.2)$$

由此还可把 X 和 X^* 进一步抽象化：设 X 和 X^* 为两个线性空间，$\langle\cdot, \cdot\rangle$: $X^*\times X\to\mathbf{R}$ 为 $X^*\times X$ 上的双线性形式(即对于固定的 $x\in X$, $\langle\cdot, x\rangle$: $X\to\mathbf{R}$ 是 X^* 上的线性形式，对于固定的 $x^*\in X^*$, $\langle x^*, \cdot\rangle$: $X\to\mathbf{R}$ 是 X 上的线性形式)，且满足(3.4.1)和(3.4.2). 则称 X 和 X^* 构成一个*对偶系*. 在这样的定义下，X 和 X^* 的地位是完全平等的. 如果为了强调它们的平等，还可用 E、F 表示两个线性空间，B: $E\times F\to\mathbf{R}$ 表示对偶双线性形式，但对于以后的应用来说，本节讨论的结果主要是用于 X 是分离局部凸空间，X^* 是其拓扑对偶的情形. 因此，在符号上我们不作更改，以便于应用时对照. 只是在本节的讨论中，我们应记住，实际上现在的 X^* 已摆脱了它对 X 的从属地位.

本节讨论的主要问题是：对于对偶系 (X, X^*), X 上有哪些局部凸拓扑，使得其拓扑对偶刚好就是 X^*.

命题 3.4.1 设 (X, X^*) 为对偶系. 如果在 X 上定义由半范数族 $\{p_i\}_{i\in I}$ 所确定的分离局部凸拓扑，其拓扑对偶恰好是 X^*，那么

$$\forall i\in I,\ \exists A_i^*\subset X^*,\ \forall x\in X, p_i(x)=\max_{x^*\in A_i^*}\langle x^*,\ x\rangle, \quad (3.4.3)$$

且

$$\bigcup_{\lambda>0}\bigcup_{i\in I}\lambda A_i^*=X^*. \quad (3.4.4)$$

证明 定义

$$A_i^*=\{x^*\in X^* | \langle x^*,\ x\rangle\leqslant p_i(x),\ \forall x\in X\} \quad (3.4.5)$$

我们指出，对于任何 $x_0\in X$，存在 $x_0^*\in A_i^*$，使得

$$\langle x_0^*,\ x_0\rangle=p_i(x_0),$$

从而(3.4.3)成立．事实上，不妨设 $x_0\neq 0$，令

$$L=\{x\in X | x=\lambda x_0,\ \lambda\in\mathbf{R}\}.$$

$x_L^*\in L^*$，满足

$$\langle x_L^*,\ x\rangle=\langle x_L^*,\ \lambda x_0\rangle=\lambda p_i(x_0),\ \forall x=\lambda x_0\in L.$$

则当 $\lambda\geqslant 0$，有

$$\langle x_L^*,\ \lambda x_0\rangle=\lambda p_i(x_0)=p_i(\lambda x_0)=p_i(x);$$

而当 $\lambda<0$ 时，有

$$\langle x_L^*,\ \lambda x_0\rangle=\lambda p_i(x_0)=-p_i(-\lambda x_0)\leqslant p_i(\lambda x_0)=p_i(x).$$

从而 $\quad\langle x_L^*,\ x\rangle\leqslant p_i(x),\ \forall x=\lambda x_0\in L.$

由 Hahn–Banach 定理 2.2.6，存在 X 上的线性形式 x_0^*，满足

$$\forall x\in L,\ \langle x_0^*,\ x\rangle=\langle x_L^*,\ x\rangle,$$
$$\forall x\in X,\ \langle x_0^*,\ x\rangle\leqslant p_i(x).$$

后式也说明 x_0^* 对于由 $\{p_i\}_{i\in I}$ 确定的 X 的局部凸拓扑连续，故 $x_0^*\in A_i^*$，且

$$\langle x_0^*,\ x_0\rangle=\langle x_L^*,\ x_0\rangle=p_i(x_0).$$

这样，(3.4.3)成立，

为验证(3.4.4)，先注意到(3.4.3)的成立也说明

$$\max_{1\leqslant k\leqslant n}p_{i_k}(x)=\max_{x^*\in\bigcup_{k=1}^n A_{i_k}^*}\langle x^*,\ x\rangle,$$

$$\bigcup_{k=1}^n A_{i_k}^*=\{x^*\in X^* | \langle x^*,\ x\rangle\leqslant\max_{1\leqslant k\leqslant n}p_{i_k}(x),\ \forall x\in X\}.$$

从而不妨假定 $\{p_i\}_{i\in I}$ 满足下列条件：

$$\forall i_1,\ \cdots,\ i_n\in I,\ \exists i\in I,\ \forall x\in X,$$

$$p_i(x)=\max_{1\le k\le n} p_{i_k}(x).$$

从而 X 中的由 $\{p_i\}$ 确定的局部凸拓扑的零邻域基的成员都可以有下列形式：

$$\{x\in X\,|\,p_i(x)\le\varepsilon\},\ \varepsilon>0,\ i\in I. \tag{3.4.6}$$

如果(3.4.4)不成立，那么由(3.4.5)，存在 $\tilde{x}^*\in X^*$，使得

$$\forall\lambda>0,\ \forall i\in I,\ \exists x_{i\lambda}\in X,\ \langle\tilde{x}^*,\ x_{i\lambda}\rangle>\lambda p_i(x_{i\lambda}).$$

但这样一来，$V=\{x\in X\,|\,|\langle\tilde{x}^*,\ x\rangle|<1\}$ 将不可能包含任何形为(3.4.6)的集合；这是因为对于任何 $\varepsilon>0$ 和 $i\in I$，可令

$$\lambda=1/\varepsilon,\ \tilde{x}_{i\varepsilon}=\varepsilon x_{i\lambda}/p_i(x_{i\lambda})=x_{i\lambda}/\lambda p_i(x_{i\lambda}),$$

从而

$$p_i(\tilde{x}_{i\varepsilon})=\varepsilon p(x_{i\lambda})/p_i(x_{i\lambda})=\varepsilon,$$

但

$$\langle\tilde{x}^*,\ \tilde{x}_{i\varepsilon}\rangle=\frac{1}{\lambda p_i(x_{i\lambda})}\langle\tilde{x}^*,\ x_{i\lambda}\rangle>1.$$

因此 V 不是零邻域．由命题 3.2.5，$\tilde{x}^*$ 对 X 的由 $\{p_i\}$ 确定的局部凸拓扑不连续．这与假定矛盾．故(3.4.4)成立．▌

由形为(3.4.3)、(3.4.4)的 $\{p_i\}$ 所确定的 X 的局部凸拓扑称为 X 对 X^* 的极化拓扑．极化的名称是由于这时 X 的零邻域

$$U_i=\{x\in X\,|\,p_i(x)\le 1\}$$

恰好是 $A_i^*\subset X^*$ 的极化集 A_i^{*0}，即

$$A_i^{*0}=U_i=\{x\in X\,|\,\langle x^*,\ x\rangle\le 1,\ \forall x^*\in A_i^*\}.$$

由于 X 与 X^* 是地位平等的，故也可同样定义 X^* 对 X 的极化拓扑．命题 3.4.1 说明了分离局部凸空间的原拓扑总是对于其拓扑对偶的极化拓扑．但是对于任意的对偶系$(X,\ X^*)$，X 关于它对 X^* 的极化拓扑来说的拓扑对偶，并不能保证总等于 X^*．我们要寻求的就是它们仍相等的条件．

最简单的极化拓扑就是对于任何 $x^*\in X^*$，$x^*\ne 0$，取

$$\forall x\in X,\ p_{x^*}(x)=|\langle x^*,\ x\rangle|=\max\{\langle x^*,\ x\rangle,\ \langle -x^*,\ x\rangle\}.$$

由 $\{p_{x^*}(x)\}_{x^*\in X^*,x^*\ne 0}$ 所确定的 X 的拓扑称为 X 对 X^* 的弱拓扑，或 $\sigma(X,\ X^*)$-拓扑．类似地也可定义 X^* 对 X 的 $\sigma(X^*,\ X)$-拓扑，但如果 X^* 是对 X 的原拓扑的拓扑对偶，这种拓扑通常称为 X^* 的弱*拓扑，可以注意到弱*拓扑总是分离的．

命题 3.4.2 设(X, X^*)为对偶系，那么 X 对于 $\sigma(X, X^*)$-拓扑的拓扑对偶是 X^*.

证明 设 $\tilde{x}^*$ 为 X 上的对于 $\sigma(X, X^*)$-拓扑的连续线性形式，那么存在 $x_1^*, \cdots, x_n^* \in X^*$，使得

$$\forall x \in X, \ \langle \tilde{x}^*, x\rangle \leqslant \max_{1\leqslant i\leqslant n} |\langle x_i^*, x\rangle|.$$

从而 $\quad \forall x \in X, \ \max\limits_{1\leqslant i\leqslant n} \{\langle x_i^* - \tilde{x}^*, x\rangle, \langle -x_i^* - \tilde{x}^*, x\rangle\} \geqslant 0.$

由定理 2.5.1，存在不全为零的 $\alpha_1, \cdots, \alpha_n, \alpha_1', \cdots, \alpha_n' \geqslant 0$，使得

$$\forall x \in X, \ \sum_{i=1}^n \alpha_i \langle x_i^* - \tilde{x}^*, x\rangle + \sum_{i=1}^n \alpha_i' \langle -x_i^* - \tilde{x}^*, x\rangle \geqslant 0,$$

即
$$\left(\sum_{i=1}^n (\alpha_i + \alpha_i')\right) \tilde{x}^* = \sum_{i=1}^n \alpha_i x_i^* - \sum_{i=1}^n \alpha_i' x_i^*.$$

因此，
$$\tilde{x}^* \in \operatorname{lin}\{x_1^*, \cdots, x_n^*\} \subset X^*. \ \blacksquare$$

注 我们也可不利用定理 2.5.1，由

$$\langle x_i^*, x\rangle = 0 (i=1, \cdots, n \Rightarrow \langle \tilde{x}^*, x\rangle = 0,$$

来直接导得 $\qquad \tilde{x}^* \in \operatorname{lin}\{x_1^*, \cdots, x_n^*\}$. $\blacksquare$

命题 3.4.2 说明 X 对于 $\sigma(X, X^*)$-拓扑的拓扑对偶为 X^*. 同样，X^* 对于 $\sigma(X^*, X)$-拓扑的拓扑对偶为 X. 这里它们有完全对称的关系. 尤其是 § 2.3 最后谈到的极化对偶关系在这种情况下就格外精巧. 下面就对此进行深入讨论.

设(X, X^*)为对偶系. $A \subset X$. $A^* \subset X^*$.

$$A^0 = \{x^* \in X^* | \langle x^*, x\rangle \leqslant 1, \ \forall x \in A\},$$

$$A^{*0} = \{x \in X | \langle x^*, x\rangle \leqslant 1, \ \forall x^* \in A^*\}.$$

分别称为 A 和 A^* 的极化集. 下列命题是极化集的基本性质:

命题 3.4.3 i) A^0 是 X^* 的包含 0 的 $\sigma(X^*, X)$-闭凸集;

ii) $A \subset A^{00}$, $A^{000} = A^0$;

iii) 如果 $A \subset B$, 那么 $B^0 \subset A^0$;

iv) $\forall \lambda \neq 0$, $(\lambda A)^0 = \lambda^{-1} A^0$;

v) $(\bigcup\limits_{i\in I} A_i)^0 = \bigcap\limits_{i\in I} A_i^0$.

证明 i) A^0 为包含 0 的凸集是明显的. 另一方面,

$$f_x(x^*)=\langle x^*, x\rangle$$

是 X^* 上的关于 $\sigma(X^*, X)$-拓扑的连续函数．而 A^0 无非是 $\mathbf{R}$ 中的闭集 $(-\infty, 1]$ 对于 f_x 的逆象，故它是 $\sigma(X^*, X)$-闭集．

ii) $A\subset A^{00}$ 由定义可得，因而也有 $A^0\subset A^{000}$．反向包含关系由 iii) 可得

iii)～v) 都由定义可得．∎

下列定理是凸集分离定理的一项重要应用：

定理 3.4.4（Banach 双极化定理）　设 (X, X^*) 为对偶系，$A\subset X$．那么

$$A^{00}=\overline{co}(A\cup\{0\})\supset\bigcup_{\lambda\in[0,1]}\lambda A,$$

这里 $\overline{co}$ 表示集合的 $\sigma(X, X^*)$-闭凸包．

证明　由命题 3.4.3 的 i)、ii)，只需证明

$$K=\overline{co}(A\cup\{0\})\supset A^{00}.$$

设 $x\notin K$．由于 X 对于 $\sigma(X, X^*)$-拓扑是分离局部凸空间，故由定理 3.3.5，存在闭超平面使 x 与 K 强分离，即存在 $x^*\in X^*$，使得

$$\langle x^*, x\rangle>\sup_{y\in K}\langle x^*, y\rangle.$$

由于 $0\in K$，故

$$\langle x^*, x\rangle>\sup_{y\in K}\langle x^*, y\rangle\geqslant 0.$$

令
$$\delta=\frac{1}{2}(\langle x^*, x\rangle+\sup_{y\in K}\langle x^*, y\rangle),$$
$$x_1^*=x^*/\delta.$$

则
$$\langle x_1^*, x\rangle>1>\sup_{y\in K}\langle x_1^*, y\rangle.$$

从而有 $x_1^*\in A^0$，$x\notin A^{00}$．这样，$K\supset A^{00}$．∎

推论 1　设 (X, X^*) 为对偶系，$A\subset X$．那么 A 是 X^* 中的集合的极化集的充要条件为 A 是包含 0 的 $\sigma(X, X^*)$-闭凸集．∎

推论 2　设 (X, X^*) 为对偶系．$A_i\subset X$，$i\in I$，都是包含 0 的 $\sigma(X, X^*)$-闭凸集．那么

$$(\bigcap_{i\in I}A_i)^0=\overline{co}\bigcup_{i\in I}A_i^0.$$

证明 由命题3.4.3的v)和定理3.4.4,

$$(\bigcap_{i\in I} A_i)^0=(\bigcap_{i\in I} A_i^{00})^0=(\bigcup_{i\in I} A_i^0)^{00}=\overline{co}\bigcup_{i\in I} A_i^0. \blacksquare$$

现在回到在本节开始所提出的问题上来. 设(X, X^*)为对偶系. 命题3.4.1说明了, 如果对X上由半范数族$\{p_i\}_{i\in I}$确定的局部凸拓扑来说,其拓扑对偶恰好是X^*,那么存在满足(3.4.4)的集合族$\{A_i^*\}_{i\in I}\subset X^*$, 使得

$$p_i(x)=\max_{x^*\in A_i^*}\langle x^*, x\rangle, \ \forall i\in I,$$

即零邻域 $U_i=\{x\in X\,|\,p_i(x)\leqslant 1\}=A_i^{*0}, \ \forall i\in I.$

在所有这种极化拓扑中,最弱的是弱拓扑$\sigma(X, X^*)$, 因为任何这样的拓扑都必须保证X^*的任何元素x^*是连续线性形式, 从而$|\langle x^*, x\rangle|$是x的连续半范数. 可能有的最强的这种拓扑则是取A_i^*为X^*中的所有对称$\sigma(X^*, X)$-有界集, 即A_i^*除有对称性来保证p_i非负外,还必须满足

$$\max_{x^*\in A_i^*}\langle x^*, x\rangle<+\infty, \quad \forall x\in X.$$

对于X^*中所有对称$\sigma(X^*, X)$-有界集所形成的X的极化拓扑, 称为X对X^*的强拓扑,通常记作$\beta(X, X^*)$. 然而, X对于强拓扑$\beta(X, X^*)$却不能保证其拓扑对偶一定是X^*, 而可能是X的代数对偶中的更大的子集. 下列定理说明了为保证拓扑对偶是X^*, 对A_i^*还有比$\sigma(X^*, X)$-有界更高的要求.

定理3.4.5(Alaoglu-Bourbaki) 设X拓扑线性空间. X^*为其拓扑对偶. U为X的对称零邻域. 那么U在X^*中的极化集U^0是$\sigma(X^*, X)$-紧凸集.

证明 首先,由于U对称,故

$$\langle x^*, x\rangle\leqslant 1, \quad \forall x\in U$$

可导致 $$\langle x^*, x\rangle\geqslant -1, \quad \forall x\in U.$$

即 $$U^0=\{x^*\in X^*\,|\,|\langle x^*, x\rangle|\leqslant 1, \quad \forall x\in U\}.$$

这样U^0可看作函数空间

$$F=\{f: U\to[-1, 1]\}=[-1, 1]^U$$

的子集. 而 F 作为紧空间 $[-1, 1]$ 的乘积空间, 其乘积拓扑在 U^0 上的诱导拓扑恰好就是 $\sigma(X^*, X)$-拓扑(它们都等价于函数的点点收敛拓扑). 由 Тихонов 定理 3.1.2, F 是紧空间, 故 U^0 是紧空间的子集. 另一方面, 由命题 3.4.3. 的 i), U^0 又是 $\sigma(X^*, X)$-闭凸集, 从而定理成立. ▮

推论　设 X 为赋范空间, X^* 为它的拓扑对偶. 那么 X^* 的闭单位球

$$\bar{B}^*=\{x^*\in X^*|\langle x^*, x\rangle\leqslant 1,\ \forall x: \|x\|\leqslant 1\}$$

是弱*($\sigma(X^*, X)$-)紧集. ▮

定理 3.4.6(Mackey-Arens)　设 (X, X^*) 为对偶系. $\tau(X, X^*)$ 表示由 X^* 中所有对称 $\sigma(X^*, X)$-紧集所决定的 X 的极化拓扑(它称为 Mackey 拓扑). 那么 X 对于 $\tau(X, X^*)$-拓扑的对偶是 X^*, 且任何使 X 的拓扑对偶为 X^* 的局部凸拓扑都一定在 $\sigma(X, X^*)$ 和 $\tau(X, X^*)$ 之间.

证明　设 X 在 $\tau(X, X^*)$-拓扑下的对偶为 X'. 显然有 $X^*\subset X'$. 需要指出 $X'\subset X^*$. 令 $x'\in X'$, 则

$$\{x'\}^0=\{x\in X|\langle x', x\rangle\leqslant 1\}$$

是 X 的对于 $\tau(X, X^*)$-拓扑的零邻域, 因为它是 $(-\infty, 1]$ 的 x'-逆象. 于是存在 X^* 的对称 $\sigma(X^*, X)$-紧集 A^*, 使得

$$A^{*0}\subset\{x'\}^0.$$

我们不妨假定 A^* 是包含 0 的凸集, 否则, 可用 A^{*00} 来代替 A^*. 对 X^* 和 X', 分别赋以 $\sigma(X^*, X)$-和 $\sigma(X', X)$-拓扑, 则 X^* 到 X' 的嵌入是连续的. 由连续映射把紧集映成紧集, A^* 也是 X' 中的包含 0 的 $\sigma(X', X)$-紧凸集, 从而在对偶系 (X, X') 中取极化, 由命题 3.4.3 的 ii)、iii) 和定理 3.4.4, 有

$$x'\in\{x'\}^{00}\subset A^{*00}=A^*\subset X^*$$

由 x' 的任意性, 即得

$$X'=X^*.$$

定理后半部是定理 3.4.5 的推论. ▮

现在已经完全回答了本节开始所提出的问题. 结论可简述如

下：设(X, X^*)是对偶系．在X上可通过X^*中的集合系来定义X中的极化拓扑．在这些极化拓扑中最弱的是弱拓扑$\sigma(X, X^*)$；最强的是强拓扑$\beta(X, X^*)$．X对于$\sigma(X, X^*)$-拓扑的拓扑对偶刚好就是X^*，但对于$\beta(X, X^*)$-拓扑的拓扑对偶则有可能比X^*大．界于$\sigma(X, X^*)$和$\beta(X, X^*)$之间还有一种 Mackey 拓扑$\tau(X, X^*)$．它能保证X对于任何界于$\sigma(X, X^*)$和$\tau(X, X^*)$之间的拓扑的拓扑对偶一定是X^*．反之亦然．这三种拓扑的定义如下：

$\sigma(X, X^*)$是对X^*的所有对称有限点集的极化拓扑．

$\beta(X, X^*)$是对X^*的所有对称$\sigma(X^*, X)$-有界集的极化拓扑．

$\tau(X, X^*)$是对X^*的所有对称$\sigma(X^*, X)$-紧集的极化拓扑．

下面讨论由此引起的一些问题，首先提出以下命题：

命题 3.4.7 设(X, X^*)为对偶系，$K\subset X$．那么对于任何$\sigma(X, X^*)$和$\tau(X, X^*)$之间的拓扑，K的闭凸包$\overline{co}\,\overline{K}$相同．∎

实际上，由定理 3.3.5 的推论，这时总有

$$\overline{co}\,K=\{x\in X|\langle x^*, x\rangle\leqslant\sigma_K(x^*),\ \forall x^*\in X^*\}.$$

下列定理指出了对于有界集有同样的情况．

定理 3.4.8(Mackey) 设(X, X^*)为对偶系．那么在X中$\sigma(X, X^*)$-有界和$\tau(X, X^*)$-有界是等价的，从而对于任何$\sigma(X, X^*)$和$\tau(X, X^*)$之间的拓扑，X的有界集族是相同的．

证明 我们只需指出$\sigma(X, X^*)$-有界集必定$\tau(X, X^*)$-有界．设A为X的$\sigma(X, X^*)$-有界集，U为X的$\tau(X, X^*)$-零邻域．不妨假设U是对称闭凸集．则由定理 3.4.4，$U^{00}=U$．我们只需指出，存在$\lambda>0$，使得$U^0\subset\lambda A^0$，因为这时由命题 3.4.3 的 ii)～iv)，就有$A\subset A^{00}\subset\lambda U^{00}=\lambda U$．

由定义，A^0是X^*的$\beta(X^*, X)$-零邻域，而由 Alaoglu-Bourbaki 定理 3.4.5，U^0是$\sigma(X^*, X)$-紧凸集．于是问题归结为证明X^*的$\beta(X^*, X)$-零邻域能够“吸收”$\sigma(X^*, X)$-紧凸集．因此，定理将是下面讨论的命题 3.4.10 的推论．∎

我们有必要对强拓扑 $\beta(X, X^*)$ 的零邻域(注意，X 与 X^* 是平等的!)进行一番讨论. 设 (X, X^*) 为对偶系. 我们称 X^* 的对称 $\sigma(X^*, X)$-有界集的极化集为 X 的桶. 如上所述，它一定是 X 的 $\beta(X, X^*)$-零邻域. 桶的概念也可直接对局部凸空间提出. 局部凸空间 X 的子集 B 称为 X 的桶，是指 B 是 X 的吸收对称闭凸集. 下列命题说明这两个定义是一致的.

命题 3.4.9 设 X 为分离局部凸空间, X^* 为其拓扑对偶. 那么 B 是 X 的桶当且仅当 B 是 X^* 的对称 $\sigma(X^*, X)$-有界集的极化集.

证明 设 B 是 X 的桶. 则由命题 3.4.7, B 也是 $\sigma(X, X^*)$-闭的; 因此，由 Banach 双极化定理 3.4.4, $B^{00}=B$. 而由 B 对称易证 B^0 也对称，由 B 吸收，则可得 B^0 是 $\sigma(X^*, X)$-有界的，因为对于任何 $x\in X$, 存在 $\lambda>0$, 使得 $\lambda x\in B$, 从而

$$\langle x^*, x\rangle\leqslant 1/\lambda,\ \forall x^*\in B^0.$$

这样 B^0 是对称 $\sigma(X^*, X)$-有界集.

现在设 $B=A^{*0}$, 而 $A^*\subset X^*$ 为对称 $\sigma(X^*, X)$-有界集. 则由命题 3.4.3, B 是包含 0 的对称 $\sigma(X, X^*)$-闭凸集. 由命题 3.4.7, 它也是原拓扑的闭凸集. B 的吸收性可由 A^* 的 $\sigma(X^*, X)$-有界性可得(如上逆推). ∎

如果局部凸空间 X 中的任何桶都是零邻域，那么 X 称为桶形空间. 由命题 3.4.9 可得以下推论.

推论 设 X 为分离局部凸空间. X^* 为其拓扑对偶. 那么 X 为桶形空间当且仅当 X 的拓扑为强拓扑 $\beta(X, X^*)$, 且这时一定有 $\beta(X, X^*)=\tau(X, X^*)$. ∎

Mackey 定理 3.4.8 的证明现在可归结为证明如下命题:

命题 3.4.10 局部凸空间 X 中的桶可吸收任何紧凸集，即设 $B\subset X$ 为桶, $A\subset X$ 为紧凸集，那么存在 $\lambda>0$, 使得 $A\subset\lambda B$.

证明 首先我们指出，如果能够证明：存在 $x\in A$, 正整数 n 和零邻域 U, 使得

$$A\cap(x+U)\subset nB, \tag{3.4.7}$$

那么命题得证. 事实上, 这时由 $A-x$ 是包含 0 的有界凸集, 存在 $\lambda \geqslant 1$, 使得

$$A-x \subset \lambda[(A-x)\cap U] \subset \lambda(nB-x),$$

因此, 再由 B 是吸收集, 存在 μ 充分大:

$$A \subset \lambda nB-(\lambda-1)x \subset \mu B.$$

如果满足(3.4.7)的 U、n、x 不存在, 那么对于任何正整数 n、任何开零邻域 U_n 和任何点 $x_n \in A$, 有

$$\{A\cap(x_n+U_n)\}\setminus nB \neq \varnothing.$$

但由 nB 是闭集, 故 $(x_n+U_n)\setminus nB$ 是开集, 从而可要求

$$x_{n+1}+\overline{U}_{n+1} \subset (x_n+U_n)\setminus nB.$$

于是闭集列 $\quad F_n = A\cap(x_n+\overline{U}_n)\,(n=1,\ 2,\ \cdots)$

(其中 x_1、U_1 任取)满足

$$F_1 \supset F_2 \supset \cdots \supset F_n \supset \cdots,\ F_n\cap nB=\varnothing.$$

由 A 是紧集,

$$\bigcap_{n=1}^{\infty} F_n \neq \varnothing.$$

令
$$\tilde{x} \in \bigcap_{n=1}^{\infty} F_n,$$

则
$$\tilde{x} \notin \bigcup_{n=1}^{\infty} nB.$$

这与 B 是吸收集矛盾. ∎

桶和桶形空间的概念是相当重要的. 下列命题指出了桶形空间类包含了最常用的局部凸空间类.

命题 3.4.11 Fréchet 空间, 即完备局部凸赋距空间, 是桶形空间. 特别是, Banach 空间——即完备赋范空间, 是桶形空间.

证明 设 X 为 Fréchet 空间, d 表示 X 上的距离, 又设 B 为 X 的桶, 即对称吸收闭凸集. 需要证明 B 是 X 的零邻域. 为此只需指出 B 有内点 x_0, 即存在 $x_0 \in B$, $\varepsilon>0$, 使得

$$U_{x_0,\varepsilon}=\{x\in X \mid d(x,\ x_0)<\varepsilon\}\subset B.$$

这是因为 B 是对称凸集, 由此可得

$$-U_{x_0,\varepsilon}=U_{-x_0,\varepsilon}=\{x\in X \mid d(x,\ -x_0)<\varepsilon\}\subset B,$$

从而 $U_{0,\varepsilon}=\frac{1}{2}(U_{x_0,\varepsilon}+U_{-x_0,\varepsilon})=\{x\in X\,|\,d(x,0)<\varepsilon\}\subset B.$

如果 B 没有内点，那么对于任何自然数 n, nB 也没有内点. 任取 $x_1\notin B$,则由于 B 是闭集,存在 $\delta_1>0$, 使得

$$U_{x_1,\delta_1}\cap B=\varnothing.$$

这里 U_{x_1,δ_1} 为以 x_1 为中心、δ_1 为半径的球. 因为 $2B$ 没有内点, 故存在 $x_2\in U_{x_1,\delta_1}$, 使得 $x_2\notin 2B$. 同样由于 $2B$ 是闭集,又存在 $\delta_2>0$, $\delta_2\leqslant\delta_1/2$, 使得

$$U_{x_2,\delta_2}\cap 2B=\varnothing.$$

依此类推,可得到点列 $\{x_n\}\subset X$, $\{\delta_n\}\subset\mathbf{R}_+$, 满足

$$U_{x_n,\delta_n}\cap nB=\varnothing,\ x_{n+1}\in U_{x_n,\delta_n},\ \delta_n/2\geqslant\delta_{n+1}>0\ (n=1,2,\cdots).$$

但 $\{x_n\}$ 显然是 Cauchy 列,从而由 X 完备,存在

$$\tilde{x}=\lim_{n\to\infty}x_n.$$

而 $\tilde{x}$ 将满足 $\tilde{x}\notin nB(n=1,2,\cdots)$. 这与 B 是吸收集相矛盾. ▌

最后，再证明一条重要定理，它指出，对于桶形空间的拓扑对偶, Mackey 定理 3.4.8 还能进一步加强.

定理 3.4.12(Banach-Steinhaus 共鸣定理) 设 X 为分离桶形空间. X^* 是其拓扑对偶. 那么在 X^* 中, $\sigma(X^*,X)$-有界和 $\beta(X^*,X)$-有界是等价的.

证明 我们只需指出,如果 A^* 是 $\sigma(X^*,X)$-有界集,那么它必定也是 $\beta(X^*,X)$-有界集; 这就是说, 如果 K 是 X 的对称 $\sigma(X^*,X)$-有界集,那么存在 $\lambda>0$, 使得 $A^*\subset\lambda K^0$. 事实上,由 $A^*\sigma(X^*,X)$-有界,故对于任何 $x\in X$, $\{\langle x^*,x\rangle\}_{x^*\in A^*}$ 有界. 令

$$B=\{x\in X\,|\,|\langle x^*,x\rangle|\leqslant 1,\ \forall x^*\in A^*\}.\tag{3.4.8}$$

则 B 是对称 $\sigma(X,X^*)$-闭凸集. 由命题 3.4.7, 和命题 3.4.9 的推论它也是对称(强)闭凸集. 再由 $\{\langle x^*,x\rangle\}_{x^*\in A^*}$ 对任何 x 有界, 可得 B 还是吸收集. 因此, B 是 X 的桶,从而是 X 的零邻域. 这样,由 K 有界,而存在 $\lambda>0$, 使得 $K\subset\lambda B$. 另一方面,由(3.4.8), 又有 $B\subset A^{*0}$, 故由命题 3.4.3 的 ii)、iii)、iv),

$$A^*\subset A^{*00}\subset B^0\subset\lambda K^0. ▌$$

§ 3.5 凸函数的连续性和对偶性

当凸函数定义在拓扑线性空间上时，我们就可讨论凸函数的连续性.

设 X 为拓扑线性空间. $f: X\to \mathbf{R}\cup\{\pm\infty\}$ 为 X 上的凸函数.

$$\operatorname{dom} f=\{x\in X \mid f(x)<+\infty\}$$

为 f 的有效域. 关于凸函数的连续性的基本定理如下:

定理 3.5.1 设 $f: X\to \mathbf{R}\cup\{+\infty\}$ 为拓扑线性空间 X 上的真凸函数. 如果存在 $x_0\in \operatorname{int}(\operatorname{dom} f)$，使得 f 在 x_0 的邻域内上有界, 即存在零邻域 U 和常数 $C\in\mathbf{R}$，满足

$$f(x)\leqslant C,\quad \forall x\in x_0+U,$$

(简称 f 在 x_0 处上有界), 那么 f 在 $\operatorname{int}(\operatorname{dom} f)$ 中连续.

证明 不妨设 $x_0=0$, $f(x_0)=0$, U 对称. 则

$$f(x)\leqslant C,\ \forall x\in U \tag{3.5.1}$$

首先指出, f 在 $x=0$ 上连续. 事实上, 任给 $\varepsilon\in(0,1)$，当 $x\in\varepsilon U$ 时，$\pm x/\varepsilon\in U$, 从而根据 f 为凸函数和(3.5.1), 有

$$f(x)\leqslant(1-\varepsilon)f(0)+\varepsilon f(x/\varepsilon)\leqslant\varepsilon C,$$

$$0=f(0)\leqslant\frac{1}{1+\varepsilon}f(x)+\frac{\varepsilon}{1+\varepsilon}f(-x/\varepsilon)$$

或
$$f(x)\geqslant-\varepsilon f(-x/\varepsilon)\geqslant-\varepsilon C.$$

因此,
$$|f(x)|\leqslant\varepsilon C,\quad \forall x\in\varepsilon U.$$

即 f 在 $x=0$ 上连续.

现在任取 $x_1\in\operatorname{int}(\operatorname{dom} f)$. 为证明 f 在 $x=x_1$ 上连续, 只需指出 f 在 x_1 邻域内上有界. 事实上, 这时存在 $\rho>1$，使得

$$y=\rho x_1\in\operatorname{dom} f.$$

从而如图 3.1, 易证

$$\forall x\in x_1+\left(1-\frac{1}{\rho}\right)U,\ \exists z\in U,$$

$$x=\left(1-\frac{1}{\rho}\right)z+\frac{1}{\rho}y.$$

因此，由 f 的凸性，有

$$\begin{aligned}f(x)&\leqslant\left(1-\frac{1}{\rho}\right)f(z)+\frac{1}{\rho}f(y)\\&\leqslant\left(1-\frac{1}{\rho}\right)C+\frac{1}{\rho}f(y)\leqslant C_1,\end{aligned}$$

即在 x_1 的邻域

$$x_1+\left(1-\frac{1}{\rho}\right)U$$

内，f 上有界. ▌

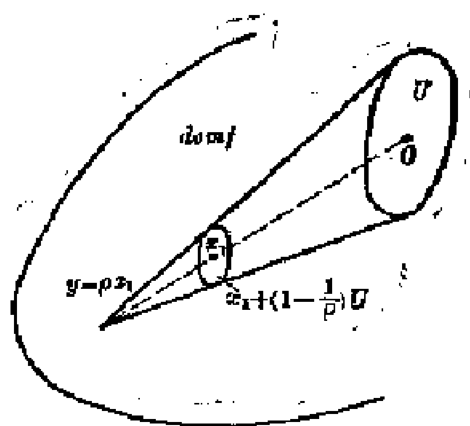


图　3.1

推论 1　如果拓扑线性空间 X 上的真凸函数 f 在 $x_0\in\operatorname{dom}f$ 上半连续，即

$$\limsup_{x\to x_0}f(x)\leqslant f(x_0),$$

那么 f 在 $\operatorname{int}(\operatorname{dom}f)$ 中连续. ▌

推论 2　$\mathbf{R}^n$ 上的真凸函数 f 在 $\operatorname{int}(\operatorname{dom}f)$ 内连续.

证明　事实上，在 $\operatorname{int}\operatorname{dom}f$ 内可找到 $n+1$ 个仿射无关的点 $a_0,\ a_1,\ \cdots,\ a_n$，则由 $a_0,\ a_1,\ \cdots,\ a_n$ 形成的单纯形 $S\subset\operatorname{int}\operatorname{dom}f$，且

$$f(x)=f\left(\sum_{i=0}^{n}\lambda_i a_i\right)\leqslant\sum_{i=0}^{n}\lambda_i f(a_i),$$

$$\forall x=\sum_{i=0}^{n}\lambda_i a_i\in S,\ \lambda_i\geqslant 0\ (i=0,\ \cdots,\ n),\ \sum_{i=0}^{n}\lambda_i=1,$$

即 f 在 S 内部上有界，而 S 的内部为 $\mathbf{R}^n$ 的开集. ▌

定理 3.5.1 指出了拓扑线性空间上的凸函数连续的条件. 尽管推论 2 指出了有限维空间上的真凸函数一定在其有效域内部连续，在无限维拓扑线性空间中，还是可能存在处处有限但处处不连续的凸函数. 例如对于任何拓扑对偶是代数对偶的真子集的拓扑线性空间来说，其上就存在不连续的线性函数(形式).

当 X 是赋范空间时，还能得到更强的结论，如下面的命题：

命题 3.5.2　设 f 为赋范空间 X 上的真凸函数. 则下列两个命题是等价的:

i) f 在 X 的一个非空开集内上有界;

ii) $\mathrm{int}(\mathrm{dom}\, f) \neq \varnothing$, 且 f 在 $\mathrm{int}(\mathrm{dom}\, f)$ 内为局部 Lipschitz 函数, 即

$$\forall x_0 \in \mathrm{int}(\mathrm{dom}\, f),\ \exists \delta_{x_0} > 0,\ \exists C_{x_0} \geqslant 0,$$
$$\forall x_1、x_2 \in B(x_0,\ \delta_{x_0}) = \{y \in X \mid \|y - x_0\| < \delta_{x_0}\},$$
$$|f(x_1) - f(x_2)| \leqslant C_{x_0} \|x_1 - x_2\|,$$

这里 $\|\cdot\|$ 表示 X 上的范数.

证明　只需证明 i) ⇒ ii). 由定理 3.5.1, f 在 $\mathrm{int}(\mathrm{dom} f)$ 内连续, 故不妨假设对于 $\delta > 0$, 有

$$|f(x)| \leqslant C,\ \forall x \in B(x_0,\ 2\delta) = \{y \in X \mid \|y - x_0\| < 2\delta\}. \tag{3.5.2}$$

我们指出,

$$|f(x_1) - f(x_2)| \leqslant \frac{2C}{\delta} \|x_2 - x_1\|,\ \forall x_1、x_2 \in B(x_0,\ \delta). \tag{3.5.3}$$

事实上, 令

$$\alpha = \|x_2 - x_1\| \neq 0,$$
$$x_3 = x_2 + (\delta/\alpha)(x_2 - x_1),$$

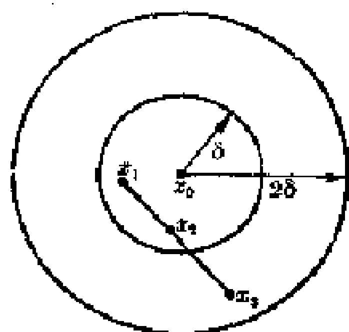


图　3.2

则 $x_3 \in B(x_0,\ 2\delta)$, 且

$$x_2 = \frac{\delta}{\alpha + \delta} x_1 + \frac{\alpha}{\alpha + \delta} x_3$$

(如图 3.2). 从而

$$f(x_2) \leqslant \frac{\delta}{\alpha + \delta} f(x_1) + \frac{\alpha}{\alpha + \delta} f(x_3),$$

$$f(x_2) - f(x_1) \leqslant \frac{\alpha}{\alpha + \delta} (f(x_3) - f(x_1))$$
$$\leqslant \frac{2\alpha}{\delta} C = \frac{2C}{\delta} \|x_2 - x_1\|,$$

即 (3.5.3) 成立. ∎

上面已经指出, 当凸函数 f 在一个点上取有限值时, 那么它在该点的连续性等价于上半连续性. 这个结果不能推广到下半连续,

例如，取 $X=\mathbf{R}$，

$$f(x)=\begin{cases}0, & \text{当 } x\leqslant 0;\\ +\infty, & \text{当 } x>0.\end{cases}$$

则 f 为凸函数，且在 $x=0$ 下半连续，即

$$\liminf_{x\to 0} f(x)\geqslant f(0).$$

但 f 在 $x=0$ 不连续．然而，如果 f 处处下半连续，那么对于一大类拓扑线性空间，仍有类似结果，如下面的命题．

命题 3.5.3　设 X 为桶形空间，$f: X\to\mathbf{R}\cup\{+\infty\}$ 为 X 上的下半连续真凸函数．那么 f 在 $\mathrm{int}(\mathrm{dom}\, f)$ 内连续．

证明　不妨设 $0\in\mathrm{int}(\mathrm{dom}\, f)$，$f(0)<a$．那么由 f 下半连续，

$$K=\{x\in X \mid f(x)\leqslant a\}$$

是非空闭凸集，且对于任何 $x_0\in X$，可取 $\varepsilon>0$ 充分小，使得

$$f\left(\frac{\varepsilon x_0}{1+\varepsilon}\right)\leqslant\frac{1}{1+\varepsilon}f(0)+\frac{\varepsilon}{1+\varepsilon}f(x_0)\leqslant a,$$

即

$$x_0\in\frac{1+\varepsilon}{\varepsilon}K.$$

因此，K 也是吸收集．这样，$U=K\cap(-K)$ 为对称吸收闭凸集，即 X 的桶．由 X 是桶形空间，U 是 X 的零邻域．于是 f 在一个零邻域内上有界．由定理 3.5.1，f 在 $\mathrm{int}(\mathrm{dom}\, f)$ 内连续．▌

推论　Fréchet 空间（包括 Banach 空间）上的下半连续真凸函数在其有效域内部连续．▌

前面我们曾经把一个凸函数 f 与它的上图 $\mathrm{epi}\, f$ 相联系，从而把许多凸函数的性质归结为图象空间中的几何性质．现在我们仍然可这样做．首先我们提出下列命题：

命题 3.5.4　设 X 为拓扑线性空间．$f: X\to\mathbf{R}\cup\{+\infty\}$ 为 X 上的真凸函数．那么

i) f 在 $\mathrm{int}(\mathrm{dom}\, f)$ 内连续等价于 $\mathrm{int}(\mathrm{epi}\, f)\neq\varnothing$；

ii) f 在 X 上下半连续等价于 $\mathrm{epi}\, f$ 是闭集．

证明　i) 因为 $\mathrm{int}(\mathrm{epi}\, f)\neq\varnothing$ 等价于 f 在某些点的邻域内

上有界，故由定理 3.5.1，它也等价于 f 在 $\mathrm{int}(\mathrm{dom}\, f)$ 内连续.

ii) f 在 X 上下半连续是指：

$$\forall x_0 \in X,\quad \liminf_{x\to x_0} f(x) \geqslant f(x_0);$$

它自然等价于：

$$(x_\alpha,\ a_\alpha) \to (x_0,\ a_0),\ f(x_\alpha) \leqslant a_\alpha,\ \forall \alpha \Rightarrow f(x_0) \leqslant a_0.$$

即 $\mathrm{epi}\, f = \{(x,\ a) \in X \times \mathbf{R} \mid f(x) \leqslant a\}$ 是闭集. ▌

由此可得到命题 2.2.4 和定理 2.2.5 对于拓扑线性空间情形的变形：

定理 3.5.5 设 f 为拓扑线性空间 X 上的真凸函数，且具有取有限值的上有界点. $A(f)$ 表示所有不大于 f 的连续仿射函数的全体，即 $a \in A(f)$ 表示存在连续的 $x_a^* \in X^*$（拓扑对偶），$\alpha_a \in \mathbf{R}$，使得

$$a(x) = \langle x_a^*,\ x\rangle + \alpha_a \leqslant f(x),\ \forall x \in X.$$

那么
$$f(x) = \sup_{a\in A(f)} a(x),\quad \forall x \in \mathrm{int}(\mathrm{dom}\, f),$$

且存在一族仿射函数 $a_i\ (i \in I)$，使得

$$f(x) = \sup_{i\in I} a_i(x),\quad \forall x \in X$$

的充要条件为 f 是 X 上的下半连续函数. ▌

此定理的证明与命题 2.2.4、定理 2.2.5 的证明类似.

值得注意的是：在这一定理中我们没有假定 X 上一定有非零连续线性形式，从而还可得到以下推论：

推论 拓扑线性空间 X 上有非零连续线性形式的充要条件为：存在取有限值的上有界点的非常数真凸函数. ▌

为了保证拓扑线性空间 X 中有足够多的连续线性形式，使上述形式的定理有意义，我们再叙述一个在分离局部凸空间中这类的定理：

定理 3.5.6 设 X 为分离局部凸空间. $f: X \to \mathbf{R} \cup \{+\infty\}$. 那么 f 为下半连续凸函数的充要条件为：它是非空连续仿射函数族 $a_i\ (i \in I)$ 的上包络：

$$f(x) = \sup_{i\in I} a_i(x),\quad \forall x \in X. ▌$$

在此应注意的是：由于可应用定理 3.3.5 及定理 3.5.6 中没有 f 在一个点的邻域内上有界的要求，因此现在不需要条件

$$\text{int}(\text{epi}\, f) \neq \varnothing.$$

推论　分离局部凸空间 X 上的下半连续真凸函数必定弱下半连续，即对于弱拓扑 $\sigma(X, X^*)$ 下半连续. ∎

这是因为连续仿射函数也一定是弱连续的，而作为弱连续函数族的上包络一定是弱下半连续的. 这个推论也可通过 epi f 是闭凸集而由命题 3.4.7 得到.

与此类似的还有以下定理.

定理 3.5.7　设 X 为分离局部凸空间. $f: X \to \mathbf{R} \cup \{+\infty\}$. 那么 f 为下半连续次线性函数的充要条件为：它是非空连续线性形式族 x_i^* $(i \in I)$ 的上包络：

$$f(x) = \sup_{i \in I} \langle x_i^*, x \rangle, \quad \forall x \in X. \ ∎$$

定理 3.5.6 和 3.5.7 等把凸函数与对偶空间中的元素联系起来. 于是如果对于对偶系 (X, X^*) 来讨论两个空间中的凸函数，我们有可能对它们建立起一种对偶关系来，从历史上看，这种对偶关系的讨论起源于力学(速度与动量之间的对偶，即所谓Legendre 变换)，并且还因此引出了对偶空间等概念. 下面对它作具体讨论：

设 (X, X^*) 为对偶系. $f: X \to \mathbf{R} \cup \{\pm\infty\}$ 为任意函数. 我们定义 $f^*: X^* \to \mathbf{R} \cup \{\pm\infty\}$ 为

$$f^*(x^*) = \sup_{x \in X} \{\langle x^*, x \rangle - f(x)\}, \quad \forall x^* \in X^*.$$

它称为 f 的共轭函数(又称极化函数). 而

$$f^{**}(x) = \sup_{x^* \in X^*} \{\langle x^*, x \rangle - f^*(x^*)\}, \quad \forall x \in X,$$

又称为 f 的二次共轭函数. 把 f 变为 f^* 的变换称为 Legendre-Fenchel 变换. 由定义得到的不等式

$$f(x) + f^*(x^*) \geqslant \langle x^*, x \rangle, \quad \forall x \in X, \ \forall x^* \in X^*,$$

称为 Young 不等式. 当 X^* 仅是 X 的拓扑对偶时，也可同样定义.

f 与 f^* 的这种对偶关系的讨论起源于力学. 为了说明这点, 我们以最简单的力学系统为例, 考虑一个质量为 m 的质点的自由落体运动. 以 x 表示质点的高度, $\dot{x}$ 表示质点的速度. 那么质点的动能为 $m\dot{x}^2/2$, 位能为 mgx, 这里 g 为重力加速度. 动能与位能的差

$$L(x, \dot{x}) = m\dot{x}^2/2 - mgx$$

称为这个质点力学系统的 Lagrange 函数. 又, 以 p 表示质点的动量, 即 $p = m\dot{x}$, 并把 p 空间看作 $\dot{x}$ 空间的对偶空间, 那么

$$\langle p, \dot{x}\rangle = p\dot{x} = m\dot{x}^2$$

的力学意义就是 2 倍的动能. 而 Lagrange 函数关于 $\dot{x}$ 的 Legendre 变换即

$$\begin{aligned} H(x, p) &= \sup_{\dot{x}} \{\langle p, \dot{x}\rangle - L(x, \dot{x})\} \\ &= \sup_{\dot{x}} \left\{p\dot{x} - \frac{m\dot{x}^2}{2} + mgx\right\} \\ &= \frac{p^2}{2m} + mgx, \end{aligned}$$

它称为系统的 Hamilton 函数, 力学意义为系统的总能量. 对于一般的力学系统也可作类似的变换. Lagrange 函数关于时间 t 的积分称为系统的作用量. 系统的运动轨线 $x = x(t)$ 是使作用量最小而决定的. 由此引出的方程称为 Euler–Lagrange 方程, 形式上它一般是 x 的二阶微分方程. 而经过 Legendre 变换后, Lagrange 函数变为 Hamilton 函数, Euler–Lagrange 方程也因此变为 Hamilton 方程, 形式上它一般是 (x, p) 的一阶微分方程组. 在下一章中将作略为深入的讨论.

f 与 f^* 的对偶关系目前在经济学中应用很多. 例如, 设 x 为产出丛, x^* 为产出价格系, $f(x)$ 为成本函数, 那么

$$f^*(x^*) = \sup_{x} \{\langle x^*, x\rangle - f(x)\}$$

称为利润函数, 即在产出价格系为 x^* 时, 企业可能获得的最大利润. 类似于成本和利润这样的对偶关系在微观经济分析中还很多.

下面重新转入数学讨论. 首先我们可注意到, 如果存在 $x_0\in X$, 使得 $f(x_0)=-\infty$, 那么将有 $f^*(x^*)\equiv+\infty$, 反之, 如果存在 $x_0^*\in X^*$, 使得 $f^*(x_0^*)=-\infty$, 那么仅当 $f(x)\equiv+\infty$ 时才有可能, 从而使 $f^*(x^*)\equiv-\infty$. 因此, 当 $f\equiv+\infty$ 或 f 可取 $-\infty$ 的情形都是没有意义的. 以后我们将只考虑 $f: X\to\mathbf{R}\cup\{+\infty\}$, 且 $f\not\equiv+\infty$ 的情形. 这时, 由定义可知 $f^*: X^*\to\mathbf{R}\cup\{+\infty\}$ 必定是 X^* 上的下半连续(对于 $\sigma(X^*, X)$ 和 $\tau(X^*, X)$ 间的任何拓扑)真凸函数. 同样, $f^{**}: X\to\mathbf{R}\cup\{+\infty\}$ 也必定是 X 上的下半连续(对于 $\sigma(X, X^*)$ 和 $\tau(X, X^*)$ 间的任何拓扑)真凸函数.

由 Young 不等式,

$$f(x)\geqslant\langle x^*, x\rangle-f^*(x^*),\ \forall x\in X,\ \forall x^*\in X^*,$$

从而
$$f(x)\geqslant\sup_{x^*\in X^*}\{\langle x^*, x\rangle-f^*(x^*)\}=f^{**}(x),\ \forall x\in X.$$

另一方面, 如果 f 是下半连续真凸函数, 则存在

$$\{x_i^*\}_{i\in I}\subset X^*,\ \{\alpha_i\}_{i\in I}\in\mathbf{R},$$

使得
$$f(x)=\sup_{i\in I}\{\langle x_i^*, x\rangle-\alpha_i\}.$$

从而
$$\begin{aligned}f^*(x^*)&=\sup_{x\in X}\{\langle x^*, x\rangle-\sup_{i\in I}(\langle x_i^*, x\rangle-\alpha_i)\}\\&=\inf_{i\in I}\sup_{x\in X}\{\langle x^*, x\rangle-\langle x_i^*, x\rangle+\alpha_i\}\end{aligned}$$

这样, 再由 Young 不等式

$$\begin{aligned}f^{**}(x)&\geqslant\langle x^*, x\rangle-f(x)\\&\geqslant\langle x^*, x\rangle-\sup_{y\in X}\{\langle x^*, y\rangle-\langle x_i^*, y\rangle+\alpha_i\}\\&\geqslant\langle x^*, x\rangle-\alpha_i,\ \forall i\in I,\ \forall x\in X.\end{aligned}$$

因此,
$$f^{**}(x)\geqslant\sup_{i\in I}\{\langle x^*, x\rangle-\alpha_i\}=f(x),\ \forall x\in X.$$

由此得到下列定理:

定理 3.5.8 f 为下半连续真凸函数当且仅当 $f=f^{**}$. ▌

在下一章中将利用凸函数的这种对偶性来表达凸规划的对偶性, 使得凸规划的对偶性有更对称的形式.

例 1 $f(x)=\langle x_0^*, x\rangle-\alpha_0$. 则

$$f^*(x^*)=\sup_{x\in X}\{\langle x^*, x\rangle-\langle x_0^*, x\rangle+\alpha_0\}$$

$$=\begin{cases}\alpha_0, & \text{当 } x^*=x_0^*;\\ +\infty, & \text{当 } x^*\neq x_0^*.\end{cases}\ ▮$$

例 2 设 $X=\mathbf{R}$, $f(x)=|x|^p/p$, $p>1$. 则

$$f^*(x^*)=|x^*|^q/q,$$

其中 q 满足 $1/p+1/q=1$. Young 不等式在这种情形即为

$$\frac{|x|^p}{p}+\frac{|x^*|^q}{q}\geqslant x^*\cdot x,\ \forall x\in\mathbf{R},\ \forall x^*\in\mathbf{R}^*,$$

它是通常所说的 Young 不等式.

例 3 设 $K\subset X$. δ_K 为其指标函数, 即

$$\delta_K(x)=\begin{cases}0, & \text{当 } x\in K;\\ +\infty, & \text{当 } x\notin K.\end{cases}$$

则

$$\delta_K^*(x^*)=\sup_{x\in X}\{\langle x^*, x\rangle-\delta_K(x)\}$$

$$=\sup_{x\in K}\langle x^*, x\rangle=\sigma_K(x^*)$$

为 K 的承托函数. 当 K 为非空闭凸集时, δ_K 为下半连续真凸函数, 从而

$$\delta_K(x)=\delta_K^{**}(x)=\sup_{x^*\in X^*}\{\langle x^*, x\rangle-\sigma_K(x^*)\}.$$

由 $x\in K$ 当且仅当 $\delta_K(x)=0$, 而 $\sigma_K(0)=0$, 我们又可得到熟知的结果(参看命题 2.3.2 和定理 3.3.5 推论)如下:

$$K=\{x\in X|\langle x^*, x\rangle\leqslant\sigma_K(x^*),\ \forall x^*\in X^*\}.$$

又, 易证 K 的极化集 K^0 为

$$K^0=\{x^*\in X^*|\langle x^*, x\rangle\leqslant 1,\ \forall x\in K\}$$

$$=\{x^*\in X^*|\sigma_K(x^*)\leqslant 1\}.$$

尽管 K^0 包含 X^* 的零, 但 σ_K 却不一定是对应 K^0 的 Minkowski 函数, 因为 σ_K 不一定非负. 但如果令

$$p_{K^0}(x^*)=\max\{0,\ \sigma_K(x^*)\},$$

则 p_{K^0} 是 X^* 上的下半连续 Minkowski 函数, 且

$$K^0=\{x^*\in X^*|p_{K^0}(x^*)\leqslant 1\}.$$

K^0 的承托函数

$$\sigma_{K^0}(x)=\sup_{x^*\in K^0}\langle x^*, x\rangle$$

则一定是 Minkowski 函数，因为 $0\in K^0$，从而 $\sigma_{K^0}(x)\geqslant 0$. 它的共轭函数显然是 K^0 的指标函数 δ_{K^0}，即

$$\delta_{K^0}(x^*)=\begin{cases}0, & \text{当 } x^*\in K^0;\\ +\infty, & \text{当 } x^*\notin K^0.\end{cases}$$

而K的双极化集 $K^{00}=\overline{co}(\{0\}\cup K)$则是

$$K^{00}=\{x\in X|\langle x^*, x\rangle\leqslant 1,\ \forall x^*\in K^0\}$$
$$=\{x\in X|\sigma_{K^0}(x)\leqslant 1\}.$$

综上所述，可得到下列结果：由 $K\subset X$ 出发，定义它的指标函数 δ_K，则它的闭凸包为

$$\overline{co}\,K=\{x\in X|\langle x^*, x\rangle\leqslant\delta_K^*(x^*),\ \forall x^*\in X^*\},$$

它的双极化包(即含 0 的闭凸包)为

$$K^{00}=\{x\in X|\sigma_{K^0}(x)\leqslant 1\},$$

其中

$$K^0=\{x^*\in X^*|\delta_K^*(x^*)\leqslant 1\},$$
$$\sigma_{K^0}(x)=\sup_{x^*\in K^0}\langle x^*, x\rangle=\delta_{K^0}^*(x).$$

当K为含 0 的闭凸集时，情况比较简单，即

$$K=\{x\in X|p_K(x)\leqslant 1\}=\{x\in X|\langle x^*, x\rangle\leqslant 1,$$
$$\forall x^*\in K^0\}=\{x\in X|\langle x^*, x\rangle\leqslant p_{K^0}(x^*),\ \forall x^*\in X^*\},$$
$$K^0=\{x^*\in X^*|p_{K^0}(x^*)\leqslant 1\}=\{x^*\in X^*|\langle x^*, x\rangle\leqslant 1,$$
$$\forall x\in K\}=\{x^*\in X^*|\langle x^*, x\rangle\leqslant p_K(x),\ \forall x\in X\},$$
$$p_K(x)=\inf\{\alpha>0|x\in\alpha K\}=\sup_{x^*\in K^0}\langle x^*, x\rangle=\delta_{K^0}^*(x),$$
$$p_{K^0}(x^*)=\inf\{\alpha>0|x^*\in\alpha K^0\}=\sup_{x\in K}\langle x^*, x\rangle=\delta_K^*(x^*).$$

当 X 为赋范空间，K 为闭单位球时，则 p_K 为 X 上的范数$\|\cdot\|_X$，K^0 为 X^* 上的闭单位球，p_{K^0} 为 X^* 上的范数$\|\cdot\|_{X^*}$，即得

$$K=\{x\in X|\|x\|\leqslant 1\}=\{x\in X|\langle x^*, x\rangle\leqslant\|x^*\|_{X^*},\ \forall x^*\in X^*\},$$
$$K^0=\{x^*\in X^*|\|x^*\|_{X^*}\leqslant 1\}=\{x^*\in X^*|\langle x^*, x\rangle\leqslant\|x\|_X,\ \forall x\in X\},$$
$$\|x\|_X=\sup_{\|X^*\|_{X^*}\leqslant 1}\langle x^*, x\rangle,$$
$$\|x^*\|_{X^*}=\sup_{\|X\|_X\leqslant 1}\langle x^*, x\rangle.$$

这些都是赋范空间理论中熟知的结果. 需注意的是 X^* 上的由 $\|\cdot\|_{X^*}$ 定义的拓扑是强拓扑 $\beta(X^*, X)$, 当它强于 Mackey 拓扑 $\tau(X^*, X)$时, X^* 的拓扑对偶将严格包含 X. 如果

$$\beta(X^*, X)=\tau(X^*, X),$$

X 称为自反空间. ▌

第三章习题

1. 试证明, 分离拓扑空间中的定向族(或滤子)至多只能有一个极限.
2. 试证明, 分离拓扑空间中的单点集是闭集, 且是该点的所有闭邻域的交.
3. 试证明, 拓扑乘积空间到每一因子空间的射影是开映射, 即开集的射影是开集. 射影是否是闭映射? 即闭集的射影是否一定是闭集?
4. 试证明, 拓扑空间 X 是分离空间的充分必要条件是: $X^2=X\times X$ 的对角线 $\Delta=\{(x, y)\in X^2|x=y\}$ 为 X^2 中的闭集.
5. 设 X 为分离拓扑空间, $\{x_n\}\subset X$ 为 X 中收敛于 $a\in X$ 的序列. 试证明, 集合 $\{a, x_1, x_2, \cdots, x_n, \cdots\}$ 是 X 的紧集. 如果把序列改为定向族, 命题是否仍成立?
6. 设 X 为紧拓扑空间. 试指出, X 中的没有聚点的子集是有限集.
7. 设 X 为紧拓扑空间, $\{x_n\}\subset X$ 为序列. 试证明, 如果 $\{x_n\}$ 有唯一的接触点 a, 那么 $\{x_n\}$ 收敛于 a. 试用反例指出, 非紧的分离拓扑空间没有这个性质.
8. 设 X 为有列紧性的距离空间. $\{O_i\}_{i\in I}$ 为 X 上的开覆盖. 试证明, 存在 $\varepsilon>0$, 使得对于任何 $x\in X$, 存在 $i\in I$, 满足
$$B(x,\varepsilon)=\{y|d(y,x)<\varepsilon\}\subset O_i.$$
(Lebesgue 引理). 由此指出, 对于距离空间来说, 紧、可数紧和列紧是等价的.
9. 设 X 为分离拓扑线性空间. $A, B\subset X$. 试证明:
   i) 如果 A 是开集, 那么 $A+B$ 也是开集.
   ii) 如果 A 是闭集, B 是紧集, 那么 $A+B$ 也是闭集; 如果 A 也是紧集, 那么 $A+B$ 也是紧集.
   iii) 试举例说明, 如果 A、B 都是闭集, $A+B$ 可能不是闭集.
10. 设 X 为线性空间. $\mathscr{U}$、$\mathscr{V}$ 分别是关于与 X 的线性结构协调的两种拓扑的零邻域基. 定义
$$\mathscr{W}=\{W\subset X|W=U\cap V,\ U\in\mathscr{U},\ V\in\mathscr{V}\},$$

试证明，以 $\mathscr{W}$ 作为零邻域基的拓扑与 X 的线性结构也协调.

11. 试证明，有限维拓扑线性空间上的任何线性形式都是连续的，并由此指出有限维线性空间上的任何分离局部凸拓扑都是相同的.

12. 试证明下列更一般的结果亦成立：任何两个 n 维分离拓扑线性空间都是拓扑同构的(代数同构是拓扑同胚).

13. 试证明，拓扑线性空间是局部紧的当且仅当它是有限维的(Riesz 定理).

14. 设 X 为线性空间，X^* 为它的代数对偶. 试证明，X 无限维当且仅当 X 上的最细局部凸拓扑不同于 $\sigma(X, X^*)$-拓扑.

15. 设 $A_1, \cdots, A_n$ 为分离拓扑线性空间 X 中的紧凸集. 试证明它们的凸包 $co\bigcup_{i=1}^{n} A_i$ 也是紧的. $\Big[$提示：令 $B=\Big\{\lambda=(\lambda_i)\in \mathbf{R}^n \,|\, \lambda_i \geqslant 0,\ i=1, \cdots, n, \sum_{i=1}^{n}\lambda_i=1\Big\}$，$f: B\times\Big(\prod_{i=1}^{n} A_i\Big)\to co\bigcup_{i=1}^{n} A_i$ 为 $f(\lambda, x_1, \cdots, x_n)=\sum_{i=1}^{n}\lambda_i x_i$. 指出 f 连续，且 $f\Big(B\times\Big(\prod_{i=1}^{n} A_i\Big)\Big)=co\bigcup_{i=1}^{n} A_i$. $\Big]$

16. 设 A、B 为分离局部凸空间 X 中的两个集合. σ_A、σ_B 分别是它们的承托函数. 试证明，$A+B$ 的承托函数为 $\sigma_{A+B}=\sigma_A+\sigma_B$，且当 A 是闭凸集，B 是紧凸集时，有

$$A+B=\{x\in X \,|\, \langle x^*, x\rangle \leqslant \sigma_A(x^*)+\sigma_B(x), \forall x^*\in X^*\}.$$

17. 设 X 为分离局部凸空间，K 为紧凸集. M 为 n 个线性无关的闭超平面的交(这里线性无关关于对应的连续线性形式而言). $K\cap M=\varnothing$. 试证明，存在 n 个线性无关的闭超平面，每一个都使 K 和 M 强分离.

18. 试证明，定理 3.3.5 中，条件"C 是 X 中的紧凸集"可代替为"C 是 X 中的 $\sigma(X, X^*)$-紧凸集".

19. 试指出，有界集的闭包、凸包、连续象以及有限代数和也是有界集；再指出，紧集是有界集.

20. 设 X 为分离局部凸空间. X^* 为其拓扑对偶. 如果 X^* 对于强拓扑 $\beta(X^*, X)$ 的拓扑对偶为 X，则 X 称为*自反空间*. 试利用定理 3.4.6 等写出自反空间的充要条件.

21. 设 X 为线性空间. X^* 为其代数对偶. 试证明 X^* 的强拓扑 $\beta(X^*, X)$ 与弱*拓扑 $\sigma(X^*, X)$ 相同.

22. 设 X 为分离局部凸空间. f、g 为 K 上的下半连续真凸函数，且

$$\operatorname{dom} f=\operatorname{dom} g=K$$

为紧凸集. 如果 $\forall x\in X$，$f(x)+g(x)>0$，试证明，存在连续线性形式 x^*，满足 $\forall x\in X, f(x)>\langle x^*, x\rangle>-g(x)$.

23. 设 X 为分离局部凸空间. f 为 X 上的真凸函数. 试证明

$$\liminf_{y\to x} f(y)=f^{**}(x); \quad \forall x\in X.$$

特别是 $$\liminf_{y\to x} f(y)\leqslant f(x), \quad \forall x\in X.$$

24．试证明共轭函数的下列性质：

i）如果 $g\leqslant f$，那么 $g^*\geqslant f^*$.

ii）如果 $g(x)=f(\lambda x)$，$\lambda\neq 0$，那么 $g^*(x^*)=f^*(x^*/\lambda)$.

iii）如果 $g(x)=\lambda f(x)$，$\lambda>0$，那么 $g^*(x^*)=\lambda f^*(x^*/\lambda)$.

iv）$(\inf\limits_{i\in I} f_i)^*=\sup\limits_{i\in I} f_i^*$，但 $(\sup\limits_{i\in I} f_i)^*\leqslant\inf\limits_{i\in I} f_i^*$.

v）如果 f、g: $X\to \mathbf{R}\cup\{+\infty\}$，那么

$$(\inf_{y\in X}\{f(y)+g(x-y)\})^*=f^*+g^*.$$

25．设 X、Y 为两个分离局部凸空间．g 为 $X\times Y$ 上的真凸函数，

$$f(x)=\inf_{y\in Y} g(x,y).$$

试证明 $$f^*(x^*)=g^*(x^*,0).$$

26．设 X、Y 为两个分离局部凸空间，F 为 $X\times Y$ 上的真凸函数，A: $X\to Y$ 为线性映射，

$$V(y)=\inf_{x\in X} F(x, Ax+y).$$

试证明 $$V^*(y^*)=F^*(-A^*y^*, y^*),$$

这里 A^*: $Y^*\to X^*$ 为 A 的共轭映射，它满足：

$$\langle Ay^*, x\rangle=\langle y^*, Ax\rangle, \quad \forall x\in X, \forall y^*\in Y^*.$$

第四章　凸函数的次微分运算

§4.1　凸函数的方向导数、Gâteaux 导数和次微分

这一章讨论凸分析的核心，即凸函数的次微分运算．关于单变量的凸函数的次微分，已经在§2.1的最后部份定义过．一般的次微分当然是它的推广．下面的讨论将在拓扑线性空间的框架中进行．其中不少内容也可按纯代数讨论，即在空间中不引进拓扑．但纯代数的情形也可看作在空间中引进了最细的局部凸拓扑，即以代数开凸吸收集全体作为零邻域基的拓扑．

设 X 为拓扑向量空间．$f: X\to\mathbf{R}\cup\{+\infty\}$ 为 X 上的真凸函数，当 $x\in\operatorname{int}(\operatorname{dom} f)$时，则对于任何 $h\in X$，$f(x+th)$是绝对值充分小的 t 的实值凸函数．于是由命题 2.1.2，

$$\delta_+ f(x;h)=\lim_{\substack{t\to 0\\ t>0}}\frac{f(x+th)-f(x)}{t}, \tag{4.1.1}$$

$$\delta_- f(x;h)=\lim_{\substack{t\to 0\\ t<0}}\frac{f(x+th)-f(x)}{t}, \tag{4.1.2}$$

都存在，且

$$\delta_+ f(x;h)\geqslant\delta_- f(x;h)=-\delta_+ f(x;-h).$$

它们分别称为 f 在 x 上沿方向 h 的右导数和左导数．如果

$$\delta_+ f(x;h)=\delta_- f(x;h), \tag{4.1.3}$$

即

$$\delta f(x;h)=\lim_{t\to 0}\frac{f(x+th)-f(x)}{t} \tag{4.1.4}$$

存在，那么 $\delta f(x;h)$称为 f 在 x 上沿方向 h 的导数．

命题 4.1.1　设 f 为拓扑线性空间 X 上的真凸函数．f 在 $x\in\operatorname{dom} f$ 上连续．那么 $\delta_+ f(x;h)$是 h 的连续次线性函数．特别

是，如果对于任何 $h\in X$，$\delta f(x;h)$都存在，那么 $\delta f(x;h)$是 h 的连续线性形式，即存在连续线性形式 $x^*\in X^*$，使得

$$\delta f(x;h)=\langle x^*,h\rangle,\ \forall h\in X.$$

证明　由(4.1.1)立即可得

$$\delta_+ f(x;\lambda h)=\lambda\delta_+ f(x;h),\ \forall\lambda\geqslant 0,\ \forall h\in X.$$

另一方面，对于任何 $h_1,h_2\in X$，由于 f 是凸函数，可得

$$\begin{aligned}\delta^+ f(x;h_1+h_2)&=\lim_{\substack{t\to 0\\ t>0}}\frac{f(x+t(h_1+h_2))-f(x)}{t}\\&\leqslant\lim_{\substack{t\to 0\\ t>0}}\frac{f(x+2th_1)+f(x+2th_2)-2f(x)}{2t}\\&=\lim_{\substack{t\to 0\\ t>0}}\frac{f(x+2th_1)-f(x)}{2t}\\&\quad+\lim_{\substack{t\to 0\\ t>0}}\frac{f(x+2th_2)-f(x)}{2t}\\&=\delta_+ f(x;h_1)+\delta_+ f(x;h_2).\end{aligned}$$

这样 $\delta_+ f(x;h)$是 h 的次线性函数.

为证明 $\delta_+ f(x;h)$ 对于任何 h 连续，由定理 3.5.1，只需证明它在一个零邻域内上有界．而对某个零邻域 U 中的 h，由 f 在 x 上连续，$f(x+h)$上有界，故

$$\delta_+ f(x;h)=\inf_{t>0}\frac{f(x+th)-f(x)}{t}\leqslant f(x+h)-f(x)$$

也上有界.

当 $\delta f(x;h)$ 对所有 h 存在时，由 $\delta f(x;h)=\delta_+ f(x;h)=-\delta_+ f(x;-h)$，可知 $\delta f(x;h)$和 $-\delta f(x;h)=\delta_+ f(x;-h)$都是 h 的连续次线性函数．因此，$\delta f(x;h)$是 h 的连续线性函数. ∎

设 $f:X\to\mathbf{R}\cup\{+\infty\}$ 为 X 上的任意函数．如果对于 $x\in\operatorname{int}(\operatorname{dom}f)$，(4.1.4)定义的 $\delta f(x;h)$对任何方向 h 存在，且存在 $x^*\in X^*$，满足

$$\delta f(x;h)=\langle x^*,h\rangle,\ \forall h\in X,$$

那么称 f 在 x 上 Gâteaux 可导，x^* 称为其 Gâteaux 导数，并记作

$x^* = \nabla f(x)$。命题 4.1.1 后半部可表达为下列推论:

推论　设 f 为拓扑线性空间 X 上的真凸函数，并且 f 在 $x \in \operatorname{dom} f$ 上连续．如果 $\delta f(x; h)$ 对任何 $h \in X$ 存在，那么 f 在 x 上 Gâteaux 可导．▌

然而，对于一般的函数来说，方向导数存在与 Gâteaux 可导是两个概念．即使对于有限维空间来说，它们也不等价．对于有限维空间 $\mathbf{R}^n$ 来说，如果 f 在 x 上 Gâteaux 可导，那么其 Gâteaux 导数即通常的梯度:

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x^1}, \cdots, \frac{\partial f(x)}{\partial x^n}\right) \in \mathbf{R}^{n*}.$$

这时，对于任何 $h = (h^1, \cdots, h^n) \in \mathbf{R}^n$，有

$$\delta f(x; h) = \lim_{t \to 0} \frac{f(x + th) - f(x)}{t} = \langle \nabla f(x), h \rangle$$

$$= \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x^i} \cdot h^i.$$

由此也容易构造方向导数存在，而 Gâteaux 微分不存在的例子．

例　我们用极坐标 (r, θ) 表示 $\mathbf{R}^2$ 中的点．令 $f: \mathbf{R}^2 \to \mathbf{R}$ 为

$$f(r, \theta) = \alpha(\theta) r,$$

其中 $\alpha(\theta)$ 为 θ 的可微函数，且满足 $\alpha(\theta) = -\alpha(\pi + \theta)$．于是 f 在原点方向导数存在．但只要例如 $\alpha(\theta)$ 不满足

$$\alpha\left(\frac{\pi}{4}\right) \neq \frac{\sqrt{2}}{2}\left(\alpha(0) + \alpha\left(\frac{\pi}{2}\right)\right) = \frac{\sqrt{2}}{2}\left(\frac{\partial f(0)}{\partial x^1} + \frac{\partial f(0)}{\partial x^2}\right),$$

f 在原点就不是 Gâteaux 可导的．▌

现在我们提出凸函数的次微分概念．设 $f: X \to \mathbf{R} \cup \{\pm\infty\}$ 为拓扑线性空间 X 上的凸函数，X^* 为 X 的拓扑对偶．下列集合称为 f 在 $x \in X$ 上的**次微分**:

$$\partial f(x) = \{x^* \in X^* \mid \langle x^*, y - x \rangle \leqslant f(y) - f(x),\ \forall y \in X\}. \tag{4.1.5}$$

这个定义实际上可不要求 f 一定是凸的，然而，对比 § 2.1 中对单变量凸函数的次微分定义可知，这个定义仅对凸函数才较有意义．

我们应当注意到，$\partial f(x)$有可能取全空间 X^*，例如 $f(x)=-\infty$，或 $f\equiv+\infty$。但当 f 为真凸函数时，这两种情况都不会出现。这时，如果 $f(x)=+\infty$，$\partial f(x)=\phi$。但即使 $f(x)\neq+\infty$，X^* 中也有足够多的元素，$\partial f(x)$ 还可能是空的。次微分中的元素称为 f 在 x 上的次梯度。如果 $\partial f(x)\neq\varnothing$，那么称 f 在 x 上次可微。$\partial f(x)\neq\varnothing$ 的 x 全体称为 ∂f 的定义域，记作 $\mathrm{dom}(\partial f)$。对于真凸函数 f，自然有

$$\mathrm{dom}(\partial f)\subset\mathrm{dom}f.$$

它们的进一步关系将在下节中讨论。此外，$\partial f(x)$显然是$\sigma(X^*, X)$-闭凸集。

定理 4.1.2 设 f 为拓扑线性空间 X 上的真凸函数，且在 $x\in\mathrm{dom}f$ 上连续。那么 $\partial f(x)\neq\varnothing$，且

$$\partial f(x)=\{x^*\in X^* \mid \langle x^*, h\rangle\leqslant\delta_+f(x;h),\ \forall h\in X\}, \tag{4.1.6}$$

$$\delta_+f(x;h)=\sup_{x^*\in\partial f(x)}\langle x^*, h\rangle,\ \forall h\in X. \tag{4.1.7}$$

证明 如果 $x^*\in X^*$ 满足

$$\langle x^*, h\rangle\leqslant\delta_+f(x;h),\ \forall h\in X$$

那么由

$$\delta_+f(x;h)=\inf_{t>0}\frac{f(x+th)-f(x)}{t}\leqslant f(x+h)-f(x)$$

可得

$$\langle x^*, h\rangle\leqslant f(x+h)-f(x),\ \forall h\in X.$$

因此，由定义(4.1.5)，$x^*\in\partial f(x)$。

反之，如果 $x^*\in\partial f(x)$，那么由定义(4.1.5)，

$$\langle x^*, h\rangle\leqslant\frac{f(x+th)-f(x)}{t},\ \forall h\in X,\ \forall t>0.$$

因此，也有

$$\langle x^*, h\rangle\leqslant\inf_{t>0}\frac{f(x+th)-f(x)}{t}=\delta_+f(x;h).$$

这样(4.1.6)成立。

$\partial f(x)\neq\phi$ 和后半部的证明可逐字照搬命题 3.4.1 的证明，这里 $\partial f(x)$ 对应 (3.4.3) 的 A_i^*，$\delta_+f(x;h)$ 对应 p_i。所不同的是

$\delta_+ f(x; h)$不是空间的半范数，从而所得到的 x_0^* 不能由半范数来说明其连续性．但现在可由定理 3.5.1，通过 x_0^* 在一个零邻域内有界来得到其连续性．∎

推论 1　设 f 为拓扑线性空间 X 上的真凸函数，并且 f 在 $x\in \mathrm{dom} f$ 上连续．那么 f 在 x 上 Gâteaux 可导当且仅当 $\partial f(x)$ 为单点集．

这由(4.1.7)可得．

推论 2　设 f 为拓扑线性空间 X 上的真凸函数，并且 f 在 $x\in \mathrm{dom} f$ 上连续．那么 $\partial f(x)$是 X^* 中的 $\sigma(X^*, X)$—紧凸集．

证明　事实上，由(4.1.7)可知，$\delta_+ f(x; h)$是 $\partial f(x)$的承托函数．由 § 3.5 的例 3 的讨论可知，

$$(\partial f(x))^0 = \{h\in X \mid \delta_+ f(x; h)\leqslant 1\}。$$

但由 $\delta_+ f(x; h)$对 h 连续，右端显然是 X 的零邻域．因此，取对称零邻域 $U\subset(\partial f(x))^0$，则由 Alaoglu-Bourbaki 定理 3.4.5，U^0 是 $\sigma(X^*, X)$–紧凸集．而 $\partial f(x)\subset(\partial f(x))^{00}\subset U^0$ 是 U^0 的 $\sigma(X^*, X)$–闭凸子集，故也是 $\sigma(X^*, X)$–紧凸集．∎

推论 3　在推论 2 的条件下，

$$co(\partial f(x)\cup\{0\}) = \bigcup_{\lambda\in[0,1]} \lambda\partial f(x)$$

是 $\sigma(X^*, X)$–紧凸集，从而有

$$(\partial f(x))^{00} = \bigcup_{\lambda\in[0,1]} \lambda\partial f(x).$$

证明　定义映射 F: $[0, 1]\times\partial f(x) \to X^*$ 为

$$F(\lambda, x^*) = \lambda x^*,\ \forall\lambda\in[0, 1],\ \forall x^*\in\partial f(x).$$

则 F 是连续映射，故作为紧空间$[0, 1]\times\partial f(x)$的连续象

$$F([0, 1]\times\partial f(x)) = \bigcup_{\lambda\in[0,1]} \lambda\partial f(x)$$

也是紧的．∎

下列定理说明凸函数的次微分与共轭函数之间的关系．

定理 4.1.3　设 f 为拓扑线性空间 X 上的真凸函数．那么下列两条件是等价的：

i)　　$x^*\in\partial f(x)$;

ii)　$$f(x)+f^*(x^*)=\langle x^*,\ x\rangle;$$

如果 $f(x)=f^{**}(x)$，特别是，如果 f 为分离局部凸空间 X 上的下半连续真凸函数，那么 i)、ii) 还等价于

iii) $x\in\partial f^*(x^*)$，这里 f^* 的次微分是把 f^* 考虑为有 $\sigma(X^*, X)$-拓扑的局部凸空间 X^* 上的凸函数来取的.

证明　如果 $x^*\in\partial f(x)$，那么

$$f(y)-f(x)\geqslant\langle x^*,\ y-x\rangle,\ \forall y\in X,$$

即

$$\langle x^*,\ x\rangle-f(x)\geqslant\langle x^*,\ y\rangle-f(y),\ \forall y\in X.$$

从而

$$\langle x^*, x\rangle-f(x)=\sup_{y\in X}\{\langle x^*,\ y\rangle-f(y)\}=f^*(x^*).$$

反之，如果 ii) 成立，由上逆推，即得 $x^*\in\partial f(x)$. 这样 i) 与 ii) 等价.

如果 $f(x)=f^{**}(x)$，则 ii) 等价于

$$f^*(x^*)+f^{**}(x)=f^*(x^*)+(f^*)^*(x)=\langle x^*,\ x\rangle,$$

以致也等价于 iii). ∎

推论 1　如果 $f(x)=f^{**}(x)$，那么 $\partial f(x)=\partial f^{**}(x)$. ∎

这是因为 $f^*=f^{***}$.

推论 2　如果 f 在 x 上次可微，即 $\partial f(x)\neq\phi$，那么 $f(x)=f^{**}(x)$. ∎

事实上，这时设 $x^*\in\partial f(x)$，则由 Young 不等式和 ii)，有

$$f^*(x^*)+f^{**}(x)\geqslant\langle x^*,\ x\rangle=f^*(x^*)+f(x).$$

因此，$f^{**}(x)\geqslant f(x)$，从而等号成立.

注　把 f 的次微分 ∂f 看作 X 到 X^* 的映射，则这是一个集值映射，即它是 X 中的点与 X^* 中的集合的对应. 对于任意两个集合 A、B，也可定义 A 到 B 的集值映射 $F: A\to 2^B$，这里 2^B 表示 B 的幂集合，即它的子集全体. 上面这样表示也就是说 A 到 B 的集值映射即 A 到 2^B 的单值映射. 乘积集合 $A\times B$ 的下列子集称为 $F: A\to 2^B$ 的图象:

$$\operatorname{graph}F=\{(x,\ y)\in A\times B\,|\,y\in F(x)\}.$$

反之，$A\times B$ 中的任何一个子集也可用来定义一个 A 到 B 的集值映射. 利用 F 的图象很容易定义 F 的逆映射 F^{-1}: $B\to 2^A$. 即只要要求 graph F^{-1}=graph F; 它也可写成:

$$y\in F(x)\Leftrightarrow x\in F^{-1}(y)$$

在这样的观点下，由定理 4.1.3 立即可得下列命题:

命题 4.1.4 设 f 为分离局部凸空间 X 上的下半连续真凸函数. 那么 f 的次微分 ∂f 和 f^* 的次微分 ∂f^* 作为集值映射互为逆映射，即

$$\partial f=(\partial f^*)^{-1},\ \partial f^*=(\partial f)^{-1}. \blacksquare$$

在下节中将进一步讨论次微分的性质. 作为这一节的结束，下面举几个次微分的具体例子.

例 1 设 X 为赋范空间. $f(x)=\|x\|$, 这里 $\|\cdot\|$ 表示 X 上的范数. 那么由 § 3.5 的例 3, 有

$$f^*(x^*)=\delta_{K^0}(x^*)=\begin{cases}0, & \text{当 } \|x^*\|\leqslant 1;\\ +\infty, & \text{当 } \|x^*\|>1.\end{cases}$$

在此我们对 X^* 上的范数用同样的符号, $K^0=\{x^*\in X^* \mid \|x^*\|\leqslant 1\}$. 因此，当 $x\neq 0$ 时，$x^*\in\partial f(x)$ 当且仅当

$$\|x^*\|\leqslant 1, \text{ 且 } \langle x^*, x\rangle=\|x\|.$$

当 $x=0$ 时，

$$\partial f(0)=\{x^*\in X^* \mid \|x^*\|\leqslant 1\}.$$

当 $X=\mathbf{R}$ 时，得到 $f(x)=|x|$ 以及

$$\partial f(x)=\begin{cases}\{1\}, & \text{当 } x>0;\\ [-1, 1], & \text{当 } x=0;\\ \{-1\}, & \text{当 } x<0.\end{cases}$$

当 $X=\mathbf{R}^n$ 时，得到 $f(x)=\left(\sum\limits_{i=1}^{n}(x^i)^2\right)^{1/2}$, 以及

$$\partial f(x)=\begin{cases}\{(x^i/\|x\|)\}, & \text{当 } x\neq 0;\\ \{x^*\in\mathbf{R}^{n*} \mid \|x^*\|\leqslant 1\}, & \text{当 } x=0.\end{cases} \qquad \blacksquare$$

例 2 仍设 X 为赋范空间. $f(x)=\dfrac{1}{2}\|x\|^2$. 则

$$f^*(x^*)=\sup_{x\in X}\left\{\langle x^*,\ x\rangle-\frac{1}{2}\|x\|^2\right\} \tag{4.1.8}$$

由于

$$\langle x^*,\ x\rangle\leqslant\|x^*\|\cdot\|x\|\leqslant\frac{1}{2}(\|x^*\|^2+\|x\|^2),$$

故

$$\langle x^*,\ x\rangle-\frac{1}{2}\|x\|^2\leqslant\frac{1}{2}\|x^*\|^2,\ \forall x\in X,\ \forall x^*\in X^*. \tag{4.1.9}$$

因为 $\|x^*\|=\sup\limits_{\|x\|\leqslant 1}\langle x^*,\ x\rangle$, 从而

$$\|x^*\|^2=\sup_{\|x\|\leqslant 1}\langle x^*,\ \|x^*\|x\rangle=\sup_{\|x\|\leqslant\|x^*\|}\langle x^*,\ x\rangle.$$

由此可得

$$\begin{aligned}\sup_{x}\left\{\langle x^*, x\rangle-\frac{1}{2}\|x\|^2\right\}&\geqslant\sup_{\|x\|\leqslant\|x^*\|}\left\{\langle x^*, x\rangle-\frac{1}{2}\|x\|^2\right\}\\&\geqslant\frac{1}{2}\|x^*\|^2.\end{aligned} \tag{4.1.10}$$

这样,由(4.1.8)～(4.1.10)得到 $f^*(x^*)=\dfrac{1}{2}\|x^*\|^2$. 而 $x^*\in\partial f(x)$ 则等价于

$$\langle x^*,\ x\rangle=\frac{1}{2}(\|x\|^2+\|x^*\|^2)$$

或

$$\langle x^*,\ x\rangle=\|x\|^2=\|x^*\|^2.$$

这时 ∂f 称为赋范空间 X 的对偶映射. 由定理 4.1.2 的推论 1 可知,对偶映射是单值映射当且仅当 $\dfrac{1}{2}\|x\|^2$ 是 Gâteaux 可导的.这种赋范空间称为光滑空间. ∎

例 3 设 K 为拓扑线性空间 X 中的凸集. δ_K 为 K 的指标函数. 那么对于任何 $x\in K$,

$$\begin{aligned}\partial\delta_K(x)&=\{x^*\in X^*|\langle x^*,\ y-x\rangle\leqslant\delta_K(y)-\delta_K(x),\ \forall y\in X\}\\&=\{x^*\in X^*|\langle x^*,\ y-x\rangle\leqslant 0,\ \forall y\in K\}.\end{aligned} \tag{4.1.11}$$

(4.1.11)右端是 X^* 中的一个 $\sigma(X^*,\ X)$-闭凸锥. 它称为 K 在 x 上的法向锥,并记作 $N_K(x)$. 当 K 是线性子空间时,

$$\partial\delta_K(x)=\{x^*\in X^*\mid\langle x^*,\ y\rangle=0,\ \forall y\in K\};$$

即 $\partial\delta_K(x)=N_K(x)$ 是 K 的"正交子空间"(在有限维时，这有完全确切的意义)，它是子空间 K 的"正交向量"全体. 当 K 是仿射集时，$N_K(x)$ 仍是类似意义的子空间. 当 K 是凸锥时，

$$\begin{aligned}\partial\delta_K(0)&=\{x^*\in X^*\mid\langle x^*,\ y\rangle\leqslant 0,\forall y\in K\}\\&=\{x^*\in X^*\mid\langle x^*,\ y\rangle\leqslant 1,\forall y\in K\}\\&=K^0.\end{aligned}$$

它自然称为 K 的极化锥或负对偶锥.

法向锥 $N_K(x)$ 在 X 中的极化锥称为 K 在 x 上的切向锥，记作 $T_K(x)$，即

$$T_K(x)=\{x\in X\mid\langle x^*,\ x\rangle\leqslant 0,\ \forall x^*\in N_K(x)\}.$$

法向锥与切向锥是两个重要的"几何"概念. 在§4.3 中要专门讨论它们. ▌

§4.2 次微分的性质

本节讨论次微分的性质. 下列基本性质由定义即可得.

命题 4.2.1 i) $\forall\lambda\geqslant 0$, $\partial(\lambda f)(x)=\lambda\partial f(x)$;

ii) $\partial f_1(x)+\partial f_2(x)\subset\partial(f_1+f_2)(x)$;

iii) 如果 $h(x)=\langle x_0^*,\ x\rangle+\alpha$, $x_0^*\in X^*$ 为 X 上的连续仿射函数，那么 $\partial h(x)=\{x_0^*\}$, 且对于 X 上的任何凸函数 f, 有

$$\partial(f+h)(x)=\partial f(x)+\partial h(x);$$

iv) f 在 x_0 上达到最小值的充要条件为 $0\in\partial f(x_0)$. ▌

定理 4.2.2 设 f_1, $\cdots$, f_n 都是拓扑线性空间 X 上的凸函数. $f(x)=\max\limits_{1\leqslant i\leqslant n}f_i(x)$, $I(x)=\{i\mid f_i(x)=f(x)\}$. 如果 f_1, $\cdots$, f_n 都在 x_0 上取有限值且连续，那么

$$\partial f(x_0)=co\bigcup_{i\in I(x_0)}\partial f_i(x_0).\tag{4.2.1}$$

证明 由于 f_1, $\cdots$, f_n 都在 x_0 上连续，故在 x_0 的某邻域上将始终有 $f(x)=\max\limits_{i\in I(x_0)}f_i(x)$. 这时，由定理 4.1.2, 首先有

$$\partial f_i(x_0)=\{x^*\in X^*\mid\langle x^*,\ h\rangle\leqslant\delta_+ f_i(x_0; h),\ \forall h\in X\}$$
$$(i=1,\ \cdots,\ m);\tag{4.2.2}$$

$$\partial f(x_0)=\{x^*\in X^*\mid\langle x^*,\ h\rangle\leqslant\delta_+ f(x_0; h),\ \forall h\in X\}.\tag{4.2.3}$$

我们先证明

$$\delta_+ f(x_0; h)=\max_{i\in I(x_0)}\delta_+ f_i(x_0; h).\tag{4.2.4}$$

事实上，当 $i\in I(x_0)$时，

$$\begin{aligned}\delta_+ f_i(x_0; h)&=\inf_{t>0}\frac{f_i(x_0+th)-f_i(x_0)}{t}\\&\leqslant\inf_{t>0}\frac{f(x_0+th)-f(x_0)}{t}\\&=\delta_+ f(x_0; h).\end{aligned}$$

故

$$\max_{i\in I(x_0)}\delta_+ f_i(x_0; h)\leqslant\delta_+ f(x_0; h).$$

另一方面，设 $t_k>0(k=1,\ 2,\ \cdots)$满足 $t_k\to 0$，且

$$\delta_+ f(x_0;h)=\lim_{k\to\infty}\frac{f(x_0+t_kh)-f(x_0)}{t_k}.$$

考虑到在 x_0 的邻域，$f(x)=\max\limits_{i\in I(X_0)} f_i(x_0)$， 所以至少存在一个 $i\in I(x_0)$，使得对于无限多个 k 满足

$$f(x_0+t_kh)=f_i(x_0+t_kh).$$

从而对于这个 i 有

$$\delta_+ f(x_0; h)=\lim_{k\to\infty}\frac{f_i(x_0+t_kh)-f_i(x_0)}{t_k}=\delta_+ f_i(x_0; h),$$

即(4.2.4)得证.

由(4.2.2)～(4.2.4)易证

$$co\bigcup_{i\in I(x_0)}\partial f_i(x_0)\subset\partial f(x_0).\tag{4.2.5}$$

为证明(4.2.5)的反向的包含式，首先注意到上式左端作为有限个 $\sigma(X^*,\ X)$-紧凸集的凸包也是 $\sigma(X^*,\ X)$-紧的(参看第三章习题 15). 另一方面，由(4.2.4)和(4.1.7)，得

$$\delta_+ f(x_0; h)=\max_{i\in I(x_0)}\ \sup_{x^*\in\partial f_i(x_0)}\langle x^*,\ h\rangle$$

$$=\sup_{x^*}\{\langle x^*, h\rangle \mid x^*\in \bigcup_{i\in I(x_0)}\partial f_i(x_0)\}.$$

这说明 $\delta_+ f(x_0; h)$ 是 $\bigcup_{i\in I(x_0)}\partial f_i(x_0)$ 的承托函数. 从而由定理 3.3.5 的推论(把 X^* 看作有 $\sigma(X^*, X)$-拓扑的分离局部凸空间),

$$\partial f(x_0)=\{x^*\in X^* \mid \langle x^*, h\rangle\leqslant \delta_+ f(x_0; h), \forall h\in X\}$$
$$=\overline{co}\bigcup_{i\in I(x_0)}\partial f_i(x_0) = co\bigcup_{i\in I(x_0)}\partial f_i(x_0).$$

即(4.2.1)得证. ▌

下列定理是次微分运算的基本定理:

定理 4.2.3(Moreau–Rockafellar) 设 X 为拓扑线性空间, f_1、f_2 为 X 上的真凸函数. 如果存在 $x_0\in(\mathrm{dom} f_1)\cap(\mathrm{dom} f_2)$, 使得 f_1 在 x_0 上连续, 那么

$$\partial(f_1+f_2)(x)=\partial f_1(x)+\partial f_2(x), \forall x\in X. \tag{4.2.6}$$

证明 由命题 4.2.1 ii), 只需证明 $\partial(f_1+f_2)(x)\subset\partial f_1(x)+\partial f_2(x)$. 设 $x^*\in\partial(f_1+f_2)(x)$, 现证明在定理条件下, $x^*\in\partial f_1(x)+\partial f_2(x)$. 不妨假设 $x^*=0$, $x=0$, $f_1(0)=f_2(0)=0$, 否则, 令

$$g_1(y)=f_1(x+y)-f_1(x)-\langle x^*, y\rangle;$$
$$g_2(y)=f_2(x+y)-f_2(x).$$

并对 g_1 和 g_2 来讨论. 这样, 可假设 $0\in\partial(f_1+f_2)(0)$, 它等价于

$$\inf_{x\in X}\{f_1(x)+f_2(x)\}=f_1(0)+f_2(0)=0. \tag{4.2.7}$$

需要指出,

$$0\in\partial f_1(0)+\partial f_2(0). \tag{4.2.8}$$

令

$$C_1=\mathrm{epi} f_1=\{(x, \alpha)\in X\times\mathbf{R} \mid f_1(x)\leqslant\alpha\}, \tag{4.2.9}$$
$$C_2=\{(x, \alpha)\in X\times\mathbf{R} \mid f_2(x)\leqslant-\alpha\}. \tag{4.2.10}$$

则由 $(\mathrm{dom} f_1)\cap(\mathrm{dom} f_2)\neq\varnothing$, f_1、f_2 都是真凸函数, 故 C_1、C_2 都是非空凸集, 且由 f_1 在 x_0 上连续, 从而在 $\mathrm{int}(\mathrm{dom} f_1)$ 中连续, 以至由命题 3.5.4 的 i), $\mathrm{int}\, C_1=\mathrm{int}(\mathrm{epi} f_1)\neq\varnothing$. 另一方面, $\mathrm{int}\, C_1\cap C_2=\varnothing$. 否则, 存在 $(x_1, \alpha_1)\in\mathrm{int}\, C_1\cap C_2$, 使得

$$f_1(x_1)<\alpha_1\leqslant -f_2(x_1),$$

以至

$$0=\inf_x\{f_1(x)+f_2(x)\}\leqslant f_1(x_1)+f_2(x_1)<0,$$

与(4.2.7)矛盾.

这样一来，我们可运用凸集分离定理 3.2.8 的推论 1. 从而存在$(x_0^*, \alpha_0^*)\in X^*\times\mathbf{R}^*$，使得

$$\langle x_0^*, y\rangle+\alpha_0^*\beta<\langle x_0^*, x\rangle+\alpha_0^*\alpha, \tag{4.2.11}$$

$$\forall(y, \beta)\in\operatorname{int}C_1, \ \forall(x, \alpha)\in C_2.$$

由左端的β可任意大，可得$\alpha_0^*\leqslant 0$. 但是如果$\alpha_0^*=0$，则由于$x_0\in(\operatorname{dom}f_1)\cap(\operatorname{dom}f_2)$，且$f_1$在$x_0$上连续，故

$$(x_0, f_1(x_0)+\varepsilon)\in\operatorname{int}C_1, \ (x_0, -f_2(x_0))\in C_2,$$

以至由(4.2.11)，有

$$\langle x_0^*, x_0\rangle<\langle x_0^*, x_0\rangle.$$

这不可能. 因此，$\alpha_0^*<0$.

最后，取$x_1^*=-x_0^*/\alpha_0^*$. 则可得

$$-\langle x_1^*, y\rangle+\beta\geqslant-\langle x_1^*, x\rangle+\alpha,$$

$$\forall(y, \beta)\in C_1, \ \forall(x, \alpha)\in C_2.$$

特别是

$$-\langle x_1^*, y\rangle+f_1(y)\geqslant-\langle x_1^*, x\rangle-f_2(x),$$

$$\forall y\in\operatorname{dom}f_1, \ \forall x\in\operatorname{dom}f_2.$$

取$x=0$，得到对于任何

$$y\in\operatorname{dom}f_1, \ f_1(y)=f_1(y)-f_1(0)\geqslant\langle x_1^*, y\rangle,$$

故$x_1^*\in\partial f_1(0)$；取$y=0$，得到对于任何

$$x\in\operatorname{dom}f_2, \ f_2(x)=f_2(x)-f_2(0)\geqslant-\langle x_1^*, x\rangle,$$

故$-x_1^*\in\partial f_2(0)$. 这样，$0\in\partial f_1(0)+\partial f_2(0)$. ∎

注 1 条件"f_1在x_0上连续"不能去掉. 下面是一个反例：取$X=\mathbf{R}$，

$$f_1(x)=\begin{cases}0, & 当\ x<0;\\ 1, & 当\ x=0;\\ +\infty, & 当\ x>0,\end{cases} \quad \operatorname{dom}f_1=(-\infty, 0],$$

$$f_2(x)=\begin{cases}+\infty, & 当\ x<0;\\ 1, & 当\ x=0;\\ 0, & 当\ x>0.\end{cases}\qquad \operatorname{dom} f_2=[0,\infty).$$

$$\partial f_1(x)=\begin{cases}\{0\}, & 当\ x<0;\\ \phi, & 当\ x\geqslant 0.\end{cases}$$

$$\partial f_2(x)=\begin{cases}\phi, & 当\ x\leqslant 0;\\ \{0\}, & 当\ x>0.\end{cases}$$

但

$$(f_1+f_2)(x)=\begin{cases}+\infty, & 当\ x\neq 0;\\ 2, & 当\ x=0.\end{cases}$$

$$\partial(f_1+f_2)(x)=\begin{cases}\phi, & 当\ x\neq 0,\\ X^*, & 当\ x=0.\end{cases}$$

即

$$\partial(f_1+f_2)(0)\neq\partial f_1(0)+\partial f_2(0).\ ▮$$

注 2 如果利用定理 3.3.4,则可得

定理 4.2.3′ 设 X 为局部凸空间. f_1、f_2 为 X 上的真凸函数. 如果存在 $x_0\in(\operatorname{dom} f_1)\cap(\operatorname{dom} f_2)$, 且 f_1 在 $\overline{\operatorname{aff}(\operatorname{dom} f_1)}$ 上的限制在 x_0 上连续, f_2 在 $\overline{\operatorname{aff}(\operatorname{dom} f_2)}$ 上的限制也在 x_0 上连续, 则

$$\partial(f_1+f_2)(x)=\partial f_1(x)+\partial f_2(x),\ \forall x\in X.\ ▮$$

这是因为这样对应的 C_1、C_2 满足 $\operatorname{ri} C_1\neq\varnothing$, $\operatorname{ri} C_2\neq\varnothing$, 且

$$(\operatorname{ri} C_1)\cap(\operatorname{ri} C_2)=\varnothing.$$

注 3 如果利用线性空间的凸集分离定理,或者说考虑线性空间的最细的局部凸拓扑,则得到定理 4.2.3 的代数陈述如下:

定理 4.2.3″ 设 X 为线性空间. f_1、f_2 为 X 上的真凸函数. 如果 $(\operatorname{dom} f_1)^{i}\cap\operatorname{dom} f_2$ 或 $(\operatorname{dom} f_1)^{ri}\cap(\operatorname{dom} f_2)^{ri}\neq\varnothing$, 那么

$$\partial(f_1+f_2)(x)=\partial f_1(x)+\partial f_2(x),\ \forall x\in X.$$

这里的次微分在代数对偶 X^* 中取值. ▮

注 4 如果希望进一步改进定理 4.2.3, 则可考虑集合

$$C=C_1-C_2=\{(y,\beta)\in X\times\mathbf{R}\mid y=x_1-x_2,\ \beta=\alpha_1-\alpha_2,\ f_1(x_1)\leqslant\alpha_1,\ f_2(x_2)\leqslant-\alpha_2\}.$$

不难证明，当 $0\in(\mathrm{dom}\,f_1-\mathrm{dom}\,f_2)^{ri}$ 时，$C^{ri}\neq\phi$. 从而定理 4.2.3″ 中的条件可减弱为

$$0\in(\mathrm{dom}\,f_1-\mathrm{dom}\,f_2)^{ri}.$$

类似的拓扑改进，可利用命题 3.5.3，即利用真凸函数的下半连续性来得到其连续性．最终结果如下：

定理 4.2.3‴ 设 X 为桶形空间，f_1、f_2 为 X 上的下半连续真凸函数．如果

$$0\in\mathrm{int}(\mathrm{dom}\,f_1-\mathrm{dom}\,f_2), \tag{4.2.12}$$

那么

$$\partial(f_1+f_2)(x)=\partial f_1(x)+\partial f_2(x),\ \forall x\in X. \blacksquare$$

在这节后半部分讨论次微分映射作为集值映射的一些性质．首先我们提出拓扑空间中的集值映射的半连续性的定义．设 A、B 为两个拓扑空间．F: $A\to 2^B$ 为 A 到 B 的集值映射．如果对于 $x\in A$，有下列性质：对于任何 B 的开集 $\Omega\supset F(x)$，存在 A 的开集 $\omega\ni x$，使得对于任何 $y\in\omega$，有 $F(y)\subset\Omega$，那么称 F 在 x 上上半连续；如果对于 $x\in A$，有下列性质：对于任何 B 的开集 Ω，满足 $\Omega\cap F(x)\neq\varnothing$，存在 A 的开集 $\omega\ni x$，使得对于任何 $y\in\omega$，有 $\Omega\cap F(y)\neq\varnothing$，那么称 F 在 x 上下半连续．如果 F 在整个空间 A 的点上都上(下)半连续，那么称 F 在 A 上上(下)半连续．如果 F 在 x 上既是上半连续的也是下半连续的，则称 F 在 x 上连续．全空间上连续的集值映射也类似定义．

很明显，当 F 为单值映射时，两种半连续性都归结为通常的连续性．但在集值映射的情形，两种半连续性是互不包含的．下面两个简单例子说明了这一点．

例 1 设 $A=B=\mathbf{R}$.

$$F(x)=\begin{cases}\{0\}, & \text{当 } x\neq 0;\\ [-1,+1], & \text{当 } x=0.\end{cases}$$

则 F 是 $\mathbf{R}$ 上的上半连续映射，但在 $x=0$ 不是下半连续的．$\blacksquare$

例 2 设 $A=B=\mathbf{R}$.

$$G(x)=\begin{cases}[-1,+1], & \text{当 } x\neq 0;\\ \{0\}, & \text{当 } x=0.\end{cases}$$

则 G 是 $\mathbf{R}$ 上的下半连续映射，但在 $x=0$ 不是上半连续的. ▌

还应注意，如果对 F 和 G 构造它们的"直径函数"：$f: \mathbf{R}\to\mathbf{R}$，$g: \mathbf{R}\to\mathbf{R}$，要求它们满足对于任何 $x\in\mathbf{R}$，有 $f(x)=d(F(x))$，$g(x)=d(G(x))$，这里 $d(A)=\max\limits_{y_1,y_2\in A}|y_1-y_2|$，那么对应的 f,g 恰好在 $x=0$ 是数值函数意义下的上、下半连续函数. 这可以作为集值映射的两种半连续性的名称的来历.

定理 4.2.4 设 X 为分离拓扑线性空间. X^* 为其拓扑对偶，f 为 X 上的真凸函数，且在 $x_0\in\operatorname{dom}f$ 上连续. 那么 $\partial f: X\to 2^{X^*}$ 在 x_0 上对于 X^* 的 $\sigma(X^*, X)$-拓扑上半连续.

证明 由定理 3.5.1，f 一定在 x_0 的一个邻域 U 中连续，从而再由定理 4.1.2，可得

$$\forall x\in U,\ \partial f(x)=\{x^*\in X^*\,|\,\langle x^*, h\rangle\leqslant\delta_+f(x;h),\ \forall h\in X\}.$$

不难验证（留给读者作为练习），只要 $\delta_+f(x;h)$ 关于 x 是上半连续的，即

$$\limsup_{y\to x}\delta_+f(y;h)\leqslant\delta_+f(x;h).$$

则 ∂f 就是对于 X^* 的 $\sigma(X^*, X)$-拓扑的上半连续映射. 而

$$\delta_+f(x;\ h)=\inf_{t>0}\frac{f(x+th)-f(x)}{t}$$

在 x_0 的邻域 U 中是一族连续函数 $\{(f(x+th)-f(x))/t\}_{t>0}$ 的下确界，故它一定是上半连续的. ▌

命题 4.2.5 设 f 为分离拓扑线性空间 X 上的真凸函数. 如果 f 的次微分映射 ∂f 在 $x\in\operatorname{dom}(\partial f)$ 上对于拓扑对偶 X^* 的 $\sigma(X^*, X)$-拓扑下半连续，那么 f 在 X 上 Gâteaux 可导. ▌

这一命题是由于次微分映射有单调性，即由定义(4.1.5)可得对于任何 $x, y\in X$ 和任何 $x^*\in\partial f(x)$、$y^*\in\partial f(y)$，有

$$\langle x^*, y-x\rangle\leqslant f(y)-f(x);$$

$$\langle y^*, x-y\rangle\leqslant f(x)-f(y),$$

从而

$$\forall x、y \in x,\ \forall x^* \in \partial f(x),\ \forall y^* \in \partial f(y),$$

$$\langle y^* - x^*,\ y - x \rangle \geqslant 0.$$

用 F 代替 ∂f，如果集值映射 $F: X \to 2^{X^*}$ 对于任何 $x、y \in X$，任何 $x^* \in F(x)$，$y^* \in F(y)$ 满足上式，那么 F 称为单调映射. 命题 4.2.5 是下面更一般的命题的推论.

命题 4.2.6 设 X 为分离拓扑线性空间，X^* 为其拓扑对偶，$F: X \to 2^{X^*}$ 为 X 到 X^* 的单调映射. 如果 F 在 $x \in \operatorname{dom} F = \{x \in X \mid F(x) \neq \phi\}$ 上对于 $\sigma(X^*, X)$-拓扑下半连续，那么 $F(x)$ 是单点集.

证明 假设 $F(x)$ 不是单点集，则存在 $x_1^*、x_2^* \in F(x)$，$x_1^* \neq x_2^*$. 从而有 $x_0 \in X$，使得

$$\langle x_2^* - x_1^*,\ x_0 \rangle = \delta > 0. \tag{4.2.13}$$

考虑 x_1^* 的 $\sigma(X^*, X)$-邻域

$$\omega_1 = \{x^* \in X^* \mid \langle x^* - x_1^*,\ x_0 \rangle \mid < \delta\}. \tag{4.2.14}$$

则由于 F 在 x 下半连续，存在 x 的邻域 U，使得

$$\forall y \in U,\ F(y) \cap \omega_1 \neq \varnothing. \tag{4.2.15}$$

取 $\lambda > 0$ 充分小，使得 $x + \lambda x_0 \in U$. 则由(4.2.14)与(4.2.15)，存在 $y_1^* \in F(x + \lambda x_0) \cap \omega_1$，即 $y_1^* \in F(x + \lambda x_0)$，

$$|\langle y_1^* - x_1^*,\ x_0 \rangle| < \delta. \tag{4.2.16}$$

但由 F 的单调性，有

$$\langle y_1^* - x_2^*,\ x + \lambda x_0 - x \rangle = \lambda \langle y_1^* - x_2^*,\ x_0 \rangle \geqslant 0,$$

即

$$\langle y_1^* - x_2^*,\ x_0 \rangle \geqslant 0. \tag{4.2.17}$$

从而由(4.2.16)与(4.2.17)，有

$$\langle x_2^* - x_1^*,\ x_0 \rangle = \langle y_1^* - x_1^*,\ x_0 \rangle - \langle y_1^* - x_2^*,\ x_0 \rangle < \delta.$$

这与(4.2.13)矛盾. ∎

次微分映射实际上还有更强的单调性，即设 $x_1, \cdots, x_n \in X$，$x_1^* \in \partial f(x_1), \cdots, x_n^* \in \partial f(x_n)$，那么由定义(4.1.5)，有

$$\langle x_1^*,\ x_2 - x_1 \rangle \leqslant f(x_2) - f(x_1)$$

$$\langle x_2^*,\ x_3 - x_2 \rangle \leqslant f(x_3) - f(x_2)$$

....

$$\langle x_n^*,\ x_1-x_n\rangle\leqslant f(x_1)-f(x_n).$$

因此,

$$\langle x_1^*,\ x_2-x_1\rangle+\langle x_2^*,\ x_3-x_2\rangle+\cdots+\langle x_n^*,\ x_1-x_n\rangle\leqslant 0.$$

用 F 代替 ∂f,则对于任何 x_i, x_i^*, n 满足上式的集值映射 $F: X\to 2^{X^*}$ 称为循环单调映射. 下列命题说明,循环单调性是次微分映射的本质特性.

命题 4.2.7(Rockafellar) 设 $(X,\ X^*)$ 为对偶系; $F:\ X\to 2^{X^*}$ 为 X 到 X^* 的集值映射. 那么存在 X 上的 $(\sigma(X,\ X^*)$-$)$下半连续真凸函数 $f: X\to\mathbf{R}\cup\{+\infty\}$满足

$$\forall x\in X,\ F(x)\subset\partial f(x) \tag{4.2.18}$$

的充要条件为 F 是循环单调映射.

证明 必要性已经指出. 下面证明充分性: 不妨设 $\mathrm{dom}\,F=\{x|F(x)\neq\varnothing\}$不是空集. 对于任何 $x\in X$, 令

$$f(x)=\sup\{\langle x_n^*,\ x-x_n\rangle+\langle x_{n-1}^*,\ x_n-x_{n-1}\rangle+\cdots+\langle x_0^*,\ x_1-x_0\rangle\},$$

这里$(x_i,\ x_i^*)$满足

$$x_i^*\in F(x_i)\quad (i=0,\ 1,\ \cdots,\ n),$$

且 x_0, $x_0^*\in F(x_0)$固定,上确界对任何 n对$(x_i,\ x_i^*)$来取. 作为连续仿射函数的上确界, f 显然是下半连续凸函数, 且由 F 的循环单调性, $f(x_0)=0$, f 不取 $-\infty$, 即 f 是下半连续真凸函数. 我们指出, 对于这个 f, (4.2.18)成立.

设 $\bar{x}\in X$, $\bar{x}^*\in F(\bar{x})$; $\alpha<f(\bar{x})$. 根据 f 的定义, 可以求得 n 对点$(x_i,\ x_i^*)(i=1,\ \cdots,\ n)$满足 $x_i^*\in F(x_i)\,(i=1,\ \cdots,\ n)$且

$$\langle x_n^*,\ \bar{x}-x_n\rangle+\cdots+\langle x_0^*,\ x_1-x_0\rangle>\alpha.$$

再由 f 的定义,又有

$$f(x)\geqslant\langle\bar{x}^*,\ x-\bar{x}\rangle+\langle x_n^*,\ \bar{x}-x_n\rangle+\cdots+\langle x_0^*,\ x_1-x_0\rangle,\quad \forall x\in X.$$

从而

$$f(x)\geqslant\langle\bar{x}^*,\ x-\bar{x}\rangle+\alpha,\ \forall\alpha<f(\bar{x}),\ \forall x\in X.$$

因此,

$$f(x)-f(\bar{x})\geqslant\langle\bar{x}^*,\ x-\bar{x}\rangle,\ \forall x\in X.$$

即 $\bar{x}^*\in\partial f(\bar{x})$, 从而(4.2.18)成立. ▌

注 用 Zorn 引理容易指出, 对于每个单调映射都存在包含(在图象集合包含的意义下)它的极大单调映射, 后者是指不存在真包含它的单调映射. 对于每个循环单调映射也都存在包含它的极大循环单调映射. 命题 4.2.7 也指出, 如果 F 是极大循环单调映射, 那么它一定是下半连续真凸函数的次微分. 这个命题的逆是否成立, 还不清楚. Rockafellar (Pacific J. of Math. 33(1970), 209~216)指出, 当 X 是 Banach 空间, 这是成立的, 甚至还指出, ∂f 是极大单调映射. ▌

§4.3 法向锥和切向锥

在前面曾提出了法向锥和切向锥的概念. 本节将对它们作进一步讨论.

设 K 为拓扑线性空间 X 中的凸集. $\delta_K(x)$ 为 K 的指标函数, 当 $x\in K$ 时,

$$N_K(x)=\partial\delta_K(x)=\{x^*\in X^*|\langle x^*,\ y-x\rangle\leqslant 0,\ \forall y\in K\}$$

称为 K 在 x 的法向锥. 当 X 为有限维并把对偶积 $\langle\cdot,\ \cdot\rangle$ 理解为数量积时, $N_K(x)$ 就是与所有向量 $y-x$ $(y\in K)$ 的夹角不小于 $\pi/2$ 的向量全体(如图 4.1).

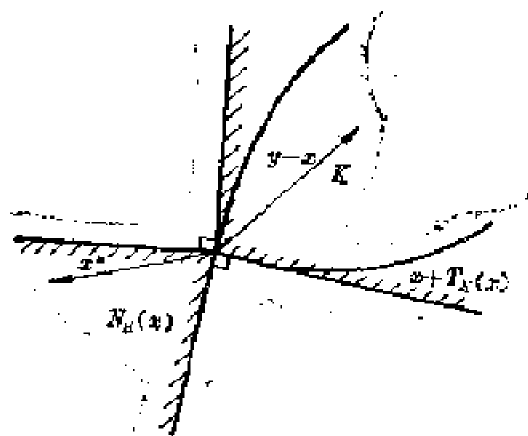


图 4.1

法向锥 $N_K(x)$ 的极化锥 $T_K(x)$ 称为 K 在 x 上的切向锥, 即

$$T_K(x)=\{x\in X|\langle x^*,\ x\rangle\leqslant 0,\ \forall x^*\in N_K(x)\}.$$

尽管对于闭凸锥来说, 上式中 $\leqslant 0$ 与 $\leqslant 1$ 是等价的, 但为了明确起见, 我们用“−”号来代替极化记号“0”, 并对其他集合也照此处理, 这

样

$$T_K(x)=N_K(x)^-. \tag{4.3.1}$$

当(X, X^*)为对偶系时，由 Banach 双极化定理 3.4.4，也有

$$N_K(x)=T_K(x)^-. \tag{4.3.2}$$

命题 4.3.1 设 K 为分离局部凸空间 X 中的凸集. 那么

$$T_K(x)=\overline{\bigcup_{\lambda>0}\lambda(K-x)}. \tag{4.3.3}$$

证明 令(4.3.3)右端为 $T(x)$. 由于 $T(x)$是闭凸锥，故由 Banach 双极化定理 $T(x)=T(x)^{--}$. 我们先证明 $N_K(x)^-\subset T(x)$. 由 $N_K(x)$也是 $(\sigma(X^*, X)^-)$闭凸锥，它等价于 $T(x)^-\subset N_K(x)$. 事实上，当 $y\in K$ 时，有 $v=y-x\in T(x)$；从而当 $x^*\in T(x)^-$ 时，有$\langle x^*, v\rangle=\langle x^*, y-x\rangle\leqslant 0$ 对所有 $y\in K$ 成立，即 $x^*\in N_K(x)$. 这样，$T(x)^-\subset N_K(x)$，从而 $N_K(x)^-\subset T(x)$.

再证明 $T(x)\subset N_K(x)^-$. 事实上，如果 $y\in T(x)$，那么存在定向族$\{\lambda_\alpha\}$, $\lambda_\alpha>0$, $\{x_\alpha\}\subset K$，使得

$$y=\lim_\alpha\lambda_\alpha(x_\alpha-x).$$

而对于任何 $x^*\in N_K(x)$，有

$$\forall\alpha,\quad \langle x^*, \lambda_\alpha(x_\alpha-x)\rangle\leqslant 0.$$

两端取极限，即得 $\langle x^*, y\rangle\leqslant 0$ 对于任何 $x^*\in N_K(x)$ 成立. 因此 $y\in N_K(x)^-$，即 $T(x)\subset N_K(x)^-$. 这就证明了

$$T(x)=N_K(x)^-=T_K(x). \blacksquare$$

命题 4.3.1 说明了切向锥 $T_K(x)$即为把原点移到 x 后，由 k 所生成的闭凸锥. 在三维情形，可把它设想为是一个能紧扣在 k 上的"帐篷"(包括内部). 当K在 x 是"光滑"的，这个"帐篷"就张平为一个切平面，这时，$T_K(x)$则是由这个切平面决定的半空间.

命题 4.3.2 设 K_1、K_2 为分离局部凸空间X 中的两个凸集. 那么有

i) 如果 $x\in K_1\subset K_2$，那么 $T_{K_1}(x)\subset T_{K_2}(x)$, $N_{K_2}(x)\subset N_{K_1}(x)$.

ii) 如果 $x_1\in K_1$, $x_2\in K_2$，那么

$$T_{K_1+K_2}(x_1+x_2)=\overline{T_{K_1}(x_1)+T_{K_2}(x_2)},$$

$$N_{K_1+K_2}(x_1+x_2)=N_{K_1}(x_1)\cap N_{K_2}(x_2).$$

iii) 如果 $x\in K_1\cap K_2$, $\operatorname{int}K_1\cap K_2$ 或 $\operatorname{ri}K_1\cap\operatorname{ri}K_2\neq\varnothing$, 那么

$$T_{K_1\cap K_2}(x)=T_{K_1}(x)\cap T_{K_2}(x),$$

$$N_{K_1\cap K_2}(x)=N_{K_1}(x)+N_{K_2}(x).$$

证明 i) 由(4.3.3)和(4.3.2)即得.

ii) 由(4.3.3),

$$\begin{aligned}T_{K_1+K_2}(x_1+x_2)&=\overline{\bigcup_{\lambda>0}\lambda(K_1+K_2-x_1-x_2)}\\&=\overline{\bigcup_{\lambda>0}\{\lambda(K_1-x_1)+\lambda(K_2-x_2)\}}\\&\subset\overline{\overline{\bigcup_{\lambda>0}\lambda(K_1-x_1)}+\overline{\bigcup_{\lambda>0}\lambda(K_2-x_2)}}\\&=\overline{T_{K_1}(x_1)+T_{K_2}(x_2)}.\end{aligned}$$

反之，由 K_1、K_2 是凸集，$x_1\in K_1$, $x_2\in K_2$, 易证

$$\bigcup_{\lambda>0}\lambda(K_1-x_1)+\bigcup_{\lambda>0}\lambda(K_2-x_2)=\bigcup_{\lambda>0}\lambda(K_1+K_2-x_1-x_2),$$

从而

$$\overline{\bigcup_{\lambda>0}\lambda(K_1-x_1)}+\overline{\bigcup_{\lambda>0}\lambda(K_2-x_2)}\subset\overline{\bigcup_{\lambda>0}\lambda(K_1+K_2-x_1-x_2)},$$

即

$$T_{K_1}(x_1)+T_{K_2}(x_2)\subset T_{K_1+K_2}(x_1+x_2).$$

从而由 $T_{K_1+K_2}(x_1+x_2)$ 是闭凸锥而得到

$$\overline{T_{K_1}(x_1)+T_{K_2}(x_2)}\subset T_{K_1+K_2}(x_1+x_2).$$

另一方面，考虑到 $x_1\in K_1$ 和 $x_2\in K_2$, 有

$$\begin{aligned}N_{K_1+K_2}(x_1+x_2)&=\{x^*\in X^*\,|\,\langle x^*,\ y-x_1-x_2\rangle\leqslant 0,\\&\qquad\forall y\in K_1+K_2\}\\&=\{x^*\in X^*\,|\,\langle x^*,\ y_1-x_1\rangle+\langle x^*,\\&\qquad y_2-x_2\rangle\leqslant 0,\ \forall y_1\in K_1,\ \forall y_2\in K_2\}\\&=\{x^*\in X^*\,|\,\langle x^*,\ y_1-x_1\rangle\leqslant 0,\\&\qquad\forall y_1\in K_1\}\cap\{x^*\in X^*\,|\,\langle x^*,\\&\qquad y_2-x_2\rangle\leqslant 0,\ \forall y_2\in K_2\}\\&=N_{K_1}(x_1)\cap N_{K_2}(x_2).\end{aligned}$$

iii) 利用 Moreau-Rockafellar 定理 4.2.3 及其变形以及

$\delta_K(x)$在x上有限连续等价于$x\in\text{int}K$等，即由命题条件可得

$$N_{K_1\cap K_2}(x)=\partial\delta_{K_1\cap K_2}(x)=\partial(\delta_{K_1}(x)+\delta_{K_2}(x))$$
$$=\partial\delta_{K_1}(x)+\partial\delta_{K_2}(x)=N_{K_1}(x)+N_{K_2}(x).$$

前者则由(4.3.1)利用上式和类似 ii) 中的讨论可得到. ▌

注 1 设C_1、C_2为闭凸锥. 那么可以证明（留给读者作为练习）:

$$(C_1\cap C_2)^-=\overline{C_1^-+C_2^-},\ (C_1+C_2)^-=C_1^-\cap C_2^-. \tag{4.3.4}$$

利用这样的关系，可使 ii)、iii) 的证明略为简化.

注 2 利用 Moreau-Rockafellar 定理 4.2.3 的改进，也可改进 iii). 例如，设X为桶形空间，K_1、K_2为闭凸集，$0\in\text{int}(K_1-K_2)$. 则对应的关系式仍成立. ▌

形为(4.3.4)的关系式有很多应用. 首先有

命题 4.3.3 设K_0, K_1, $\cdots$, K_n为分离局部凸空间X中的$n+1$个凸锥，且

$$K_0\cap\text{int}\,K_1\cap\cdots\cap\text{int}\,K_n$$

或
$$\text{ri}K_0\cap\text{ri}K_1\cap\cdots\cap\text{ri}K_n\neq\varnothing. \tag{4.3.5}$$

那么

$$\left(\bigcap_{i=0}^n K_i\right)^-=\sum_{i=0}^n K_i^-. \tag{4.3.6}$$

证明 不妨设$0\in K_i(i=0,1,\cdots,n)$. 则

$$K_i^-=\{x^*\in X^*|\langle x^*,x\rangle\leqslant 0,\ \forall x\in K_i\}$$
$$=\partial\delta_{K_i}(0).$$

于是(4.3.6)等价于

$$\partial\delta_{\sum\limits_{i=0}^n K_i}(0)=\partial\left(\sum_{i=0}^n\delta_{K_i}(0)\right)=\sum_{i=0}^n\partial\delta_{K_i}(0).$$

命题归结为用(4.3.5)验证 Moreau-Rockafellar 定理 4.2.3 的条件成立. ▌

定理 4.3.4(Дубовицкий-Милютин) 设K_0, K_1, $\cdots$, K_n为分离局部凸空间X中的$n+1$个非空凸锥. 如果

$$\bigcap_{i=1}^n\text{int}K_i\neq\varnothing, \tag{4.3.7}$$

那么

$$K_0\cap\left(\bigcap_{i=1}^{n}\operatorname{int}K_i\right)=\varnothing \tag{4.3.8}$$

的充要条件为存在

$$x_i^*\in K_i^- \quad (i=0,\ 1,\ 2,\ \cdots,\ n)$$

使得

$$\langle x_0^*,\ y\rangle>0,\ \forall y\in\bigcap_{i=1}^{n}\operatorname{int}K_i, \tag{4.3.9}$$

且

$$\sum_{i=0}^{n}x_i^*=0. \tag{4.3.10}$$

证明 如果 (4.3.8) 成立，那么由 (4.3.7) 和凸集分离定理 3.2.8 的推论 1，存在连续线性形式 $x_0^*\neq 0$，使得

$$\langle x_0^*,\ x\rangle<\langle x_0^*,\ y\rangle,\ \forall x\in K_0,\ \ \forall y\in\bigcap_{i=1}^{n}\operatorname{int}K_i.$$

不妨设 $0\in K_0$，则(4.3.9)成立，且 $-x_0^*\in\left(\bigcap_{i=1}^{n}\operatorname{int}K_i\right)^-$. 由命题 4.3.3，存在 $x_i^*\in K_i^-(i=1,\ \cdots,\ n)$，使得

$$-x_0^*=\sum_{i=1}^{n}x_i^*,$$

即(4.3.10)成立. $x_0^*\in K_0^-$ 可由 K_0 是凸锥而导得.

反之，如果存在 $x_i^*\in K_i^-$ $(i=0,\ 1,\ 2,\ \cdots,\ n)$，使得(4.3.9)、(4.3.10)成立，但又存在 $x_0\in K_0\cap\left(\bigcap_{i=1}^{n}\operatorname{int}K_i\right)$，那么有

$$\langle x_0^*,\ x_0\rangle\leqslant 0,$$

$$\left\langle\sum_{i=1}^{n}x_i^*,\ x_0\right\rangle=-\langle x_0^*,\ x_0\rangle<0.$$

导致矛盾. ▌

注 把(4.3.7)、(4.3.8)代替为

$$\operatorname{ri}K_0\neq\phi,\ \bigcap_{i=1}^{n}\operatorname{ri}K_i\neq\varnothing, \tag{4.3.7$'$}$$

$$\bigcap_{i=0}^{n}\operatorname{ri}K_i=\varnothing, \tag{4.3.8$'$}$$

则定理仍成立. ▌

定理 4.3.4 有时也称为凸锥分离定理，它在变分学、规划论、最优控制等多方面都有应用．在下节我们要指出它在凸规划理论中的应用．

在本节最后，我们指出一个非常有用的结果如下：

定理 4.3.5 设 g 为分离局部凸空间 X 上的真凸函数，且存在 $x_1\in X$，满足 $g(x_1)<0$；

$$K=\{x\in X \mid g(x)\leqslant 0\}. \tag{4.3.11}$$

那么，当 g 在 x 上连续时，有

$$\partial\delta_K(x)=\begin{cases}\bigcup\limits_{\lambda>0}\lambda\partial g(x), & \text{当 } g(x)=0;\\ \{0\}, & \text{当 } g(x)<0;\\ \phi, & \text{当 } g(x)>0.\end{cases} \tag{4.3.12}$$

证明 当 $g(x)>0$ 时，$\delta_K(x)=+\infty$，故 $\partial\delta_K(x)=\varnothing$；当 $g(x)<0$ 时，则 $x\in \mathrm{int}(\mathrm{dom}\delta_K)$，故 $\partial\delta_K(x)=\{0\}$．当 $g(x)=0$ 时，如果 $x^*\in\partial g(x)$，则

$$\langle x^*,\ y-x\rangle\leqslant g(y)-g(x)=g(y),\ \forall y\in X.$$

从而由(4.3.11)，

$$\langle x^*,\ y-x\rangle\leqslant 0,\ \forall y\in K,\ \forall x^*\in\partial g(x).$$

因此，

$$T_K(x)=\overline{\bigcup_{\lambda>0}\lambda(k-x)}\subset(\partial g(x))^-.$$

这里对 $\partial g(x)$ 的"－"号，正如前面所说的，与对凸锥的"－"号意义一样．

现在我们指出 $(\partial g(x))^-\subset T_K(x)$．设 $h\in(\partial g(x))^-$．那么

$$\langle x^*,\ h\rangle\leqslant 0,\ \forall x^*\in\partial g(x),$$

从而由(4.1.7)，有

$$\delta_+g(x;h)=\sup_{x^*\in\partial g(x)}\langle x^*,\ h\rangle\leqslant 0.$$

另一方面，由 $g(x_1)<0$ 可得

$$\delta_+g(x;x_1-x)\leqslant g(x_1)-g(x)<0,$$

从而对于任何 $y_\lambda=(1-\lambda)(x_1-x)+\lambda h$，$\lambda\in[0,\ 1)$，有

$$\delta_+g(x;y_\lambda)\leqslant(1-\lambda)\delta_+g(x;x_1-x)+\lambda\delta_+g(x;h)<0,$$

即

$$\lim_{\substack{t\to 0\\ t>0}}\frac{g(x+ty_\lambda)-g(x)}{t}=\inf_{t>0}\frac{g(x+ty_\lambda)}{t}<0.$$

因此，对充分小的 $t>0$，有 $g(x+ty_\lambda)<0$，以至

$$y_\lambda\in\frac{1}{t}(K-x)\subset T_K(x).$$

由 $T_K(x)$ 是闭凸锥，故 $h=\lim\limits_{\lambda\to 1}y_\lambda\in T_K(x)$.

最后还需指出，

$$\partial\delta_K(x)=N_K(x)=T_K(x)^-=(\partial g(x))^{--}=\bigcup_{\lambda\geq 0}\lambda\partial g(x).$$

首先由 Banach 双极化定理 3.4.4，易证

$$(\partial g(x))^{--}=\overline{\bigcup_{\lambda\geq 0}\lambda\partial g(x)}.$$

这里闭包对 $\sigma(X^*, X)$-拓扑来取. 因此，只需指出 $(\partial g(x))^{--}\subset\bigcup\limits_{\lambda\geq 0}\lambda\partial g(x)$. 设 $x^*\in(\partial g(x))^{--}$，则存在定向族 $\{\lambda_\alpha\}$, $\lambda_\alpha>0$, $\{x_\alpha^*\}\subset\partial g(x)$，使得 $x_0^*=\lim\limits_\alpha\lambda_\alpha x_\alpha^*$. 我们指出 $\{\lambda_\alpha\}$ 有界. 事实上，如果 $\{\lambda_\alpha\}$ 无界，不妨设 $\lambda_\alpha\to+\infty$. 则由于 $\delta_+g(x;x_1-x)=-\delta<0$，有

$$\langle x^*, x_1-x\rangle\leqslant\delta_+g(x;x_1-x)=-\delta,\ \forall x^*\in\partial g(x).$$

从而

$$\langle\lambda_\alpha x_\alpha^*, x_1-x\rangle\leqslant-\lambda_\alpha\delta,\ \forall\alpha.$$

两端取极限得 $\langle x_0^*, x_1-x\rangle\leqslant-\infty$，这不可能. 这样，存在常数 $c>0$，使得 $\{\lambda_\alpha x_\alpha^*\}\subset\bigcup\limits_{\lambda\in[0,c]}\lambda\partial g(x)$. 由定理 4.1.2 的推论 3,

$$\bigcup_{\lambda\in[0,c]}\lambda\partial g(x)=c\bigcup_{\lambda\in[0,1]}\lambda\partial g(x)$$

是 $\sigma(X^*, X)$-紧凸集，故

$$x_0^*\in\bigcup_{\lambda\in[0,c]}\lambda\partial g(x)\subset\bigcup_{\lambda\geq 0}\lambda\partial g(x).$$

即 $(\partial g(x))^{--}\subset\bigcup\limits_{\lambda\geq 0}\lambda\partial g(x)$. ∎

注 1 当 $g(x)=0$ 时，条件"存在 $x_1\in X$，满足 $g(x_1)<0$"等价于"g 不在 x 达到最小值"，即"$0\notin\partial g(x)$". ∎

注 2 当 $g(x)=0$，且在 x Gâteaux 可导时，$\partial g(x)=\{\nabla g(x)\}$.

从而当 $\nabla g(x)\neq 0$ 时,

$$N_K(x)=\partial\delta_K(x)=\{x^*\in X^*\mid x^*=\lambda\nabla g(x),\ \lambda\geqslant 0\},$$
$$T_K(x)=\{h\in X\mid\langle\nabla g(x),\ h\rangle\leqslant 0\}.$$

这时,法向锥退化为法向量,而切向锥退化为切半空间,其边界为"凸流形"$\{x\in X\mid g(x)=0\}$在点 x 的"切空间"

$$\{h\in X\mid\langle\nabla g(x),\ h\rangle=0\}.\ ▮$$

注3　当 g 为仿射函数 $g(x)=\langle x_0^*,\ x\rangle-\alpha$ 时, $\partial g(x)=\{x_0^*\}$, 从而

$$N_K(x)=\{x^*\in X^*\mid x^*=\lambda x_0^*,\ \lambda\geqslant 0\},$$
$$T_K(x)=\{x\in X\mid\langle x_0^*,\ x\rangle\leqslant 0\}=K-x.\quad ▮$$

§4.4　凸规划问题上的应用

在这一节中先利用次微分运算来讨论前面已经讨论过的凸规划问题$(\mathscr{P})$,即

$$(\mathscr{P})\begin{cases}f(x)\to\min,\\ g_i(x)\leqslant 0,\quad i=1,\ \cdots,\ p;\\ h_j(x)=0,\quad j=1,\ \cdots,\ q.\end{cases}$$

首先我们有下列结果:

定理4.4.1(Fritz John 条件)　设 $f,\ g_1,\ \cdots,\ g_p$ 是拓扑线性空间 X 上的真凸函数, $h_1,\ \cdots,\ h_q$ 为 X 上的连续仿射函数, $f,\ g_1,\ \cdots,\ g_p$ 都在 $\hat{x}\in X$ 上有限且连续, $h(\hat{x})=(h_1(\hat{x}),\ \cdots,\ h_q(\hat{x}))=0$, 那么 $\hat{x}$ 是问题$(\mathscr{P})$的解的充分必要条件为存在 $\hat{\lambda}_0\geqslant 0,\ \hat{\lambda}_1,\ \cdots,\ \hat{\lambda}_p\geqslant 0$, 且不全为零, 以及 $\hat{\mu}_1,\ \cdots,\ \hat{\mu}_q\in\mathbf{R}$, 使得

$$0\in\hat{\lambda}_0\partial f(\hat{x})+\sum_{i=1}^{p}\hat{\lambda}_i\partial g_i(\hat{x})+\sum_{j=1}^{q}\hat{\mu}_j\partial h_j(\hat{x}),\tag{4.4.1}$$

$$\hat{\lambda}_i g_i(\hat{x})=0\quad(i=1,\ \cdots,\ p).\tag{4.4.2}$$

证明　先假设 $h=(h_1,\ \cdots,\ h_q):\ X\to\mathbf{R}^q$ 是满射. 这时因为 $f,\ g_1,\ \cdots,\ g_p$ 都在 $\hat{x}$ 上有限连续, 故

$$\hat{x}\in(\operatorname{dom}f)^i\cap\left(\bigcap_{i=1}^{p}(\operatorname{dom}g_i)^i\right).$$

同时由 $h(\hat{x})=0$ 以及 h 是仿射满射，可得

$$0=h(\hat{x})\in\left[h\left(\operatorname{dom}f\cap\left(\bigcap_{i=1}^{p}\operatorname{dom}g_i\right)\right)\right]^i.$$

与前面一样，问题 $(\mathscr{P})$ 有解 $\hat{x}$ 等价于

$$\begin{cases}F(x)=\max\limits_{1\leqslant i\leqslant p}\{f(x)-f(\hat{x}),\ g_i(x)\}\geqslant 0,\\ \forall x;\ h(x)=(h_1(x),\ \cdots,\ h_q(x))=0.\end{cases}$$

因为

$$0\in\left[h\left(\operatorname{dom}f\cap\left(\bigcap_{i=1}^{p}\operatorname{dom}g_i\right)\right)\right]^i=[h(\operatorname{dom}F)]^i,$$

故由定理 2.5.1，上式成立的充分必要条件为存在 $\alpha>0$, $\hat{\mu}_1,\ \cdots,\ \hat{\mu}_q\in\mathbf{R}$，使得

$$\alpha F(x)+\sum_{j=1}^{q}\hat{\mu}_jh_j(x)\geqslant 0,\quad\forall x\in\operatorname{dom}F.$$

但由于 $\alpha>0$，上式等价于

$$\alpha F(x)+\sum_{j=1}^{q}\hat{\mu}_jh_j(x)\geqslant 0,\quad\forall x\in X.$$

上式意味着 $\alpha F(x)+\sum\limits_{j=1}^{q}\hat{\mu}_jh_j(x)$ 在 $\hat{x}\in X$ 上达到最小值 0，故其充要条件为

$$\begin{aligned}0&\in\partial\left(\alpha F+\sum_{j=1}^{q}\hat{\mu}_jh_j\right)(\hat{x})\\&=\alpha\partial F(\hat{x})+\sum_{j=1}^{q}\hat{\mu}_j\partial h_j(\hat{x}),\end{aligned}$$

后式不利用 Moreau-Rockafellar 定理 4.2.3 也可得到（命题 4.2.1.iii)）．由定理 4.2.2，

$$\partial F(\hat{x})=co\{\partial f(\hat{x}),\ \bigcup_{i\in I(\hat{x})}\partial g_i(\hat{x})\},$$

这里 $I(\hat{x})=\{i\mid g_i(\hat{x})=0\}$．于是存在

$$\lambda_0\geqslant 0;\ \lambda_i\geqslant 0,\ \forall i\in I(\hat{x});$$

$$\lambda_i=0,\ \forall i\notin I(\hat{x});\ \sum_{i=0}^{p}\lambda_i=1,$$

使得

$$0\in\alpha\left\{\lambda_0\partial f(\hat{x})+\sum_{i=1}^{p}\lambda_i\partial g_i(\hat{x})\right\}+\sum_{j=1}^{q}\hat{\mu}_j\partial h_j(\hat{x}).$$

取 $\hat{\lambda}_0=\alpha\lambda_0$, $\hat{\lambda}_i=\alpha\lambda_i$ $(i=1, \cdots, p)$, 则(4.4.1)、(4.4.2)成立.

如果 $h: X\to \mathbf{R}^q$ 不是满射, 那么可取 $h(X)(\neq\phi)$ 代替 $\mathbf{R}^q$ 来作同样的讨论. ∎

定理 4.4.2 (Kuhn-Tucker 条件)　在定理 4.4.1 的条件下, 如果存在 $x_0\in X$, 使得下列 Slater 条件成立:

(S)　$g_i(x_0)<0 (i=1, \cdots, p)$; $h_j(x_0)=0$ $(j=1, \cdots, q)$,

那么(4.4.1)中的 $\hat{\lambda}_0=1$, 即这时问题($\mathscr{P}$)有解为 $\hat{x}$ 的充要条件为存在 $\hat{\lambda}_1, \cdots, \hat{\lambda}_p\geqslant 0$, $\hat{\mu}_1, \cdots, \hat{\mu}_q\in\mathbf{R}$, 使得

$$0\in\partial f(\hat{x})+\sum_{i=1}^{p}\hat{\lambda}_i\partial g_i(\hat{x})+\sum_{j=1}^{q}\hat{\mu}_j\partial h_j(\hat{x}), \tag{4.4.3}$$

$$\hat{\lambda}_i g_i(\hat{x})=0 \quad (i=1, \cdots, p). \tag{4.4.2}$$

证明　只需证明必要性, 如果 $\hat{\lambda}_0=0$, 那么由(4.4.1),

$$0\in\sum_{i=1}^{p}\hat{\lambda}_i\partial g_i(\hat{x})+\sum_{j=1}^{q}\hat{\mu}_j\partial h_j(\hat{x}).$$

这说明 $\sum_{i=1}^{p}\hat{\lambda}_i g_i(x)+\sum_{j=1}^{q}\hat{\mu}_j h_j(x)$ 在 $x=\hat{x}$ 上有最小值为零. 但 $\hat{\lambda}_1, \cdots$ $\hat{\lambda}_p$ 不全为零, 故它与条件(S)矛盾. ∎

比较一下定理 4.4.1、4.4.2 与定理 2.5.2、2.5.3 的区别是很有意思的. 我们增加的条件是 f、g_i 等都在 $\hat{x}$ 上连续, 且 $h(\hat{x})=0$, 换来的却是 (2.5.6) 对所有 x 都成立. 这样的条件也取代了原来的条件(S)ii). 当 f、g_i 等不取 $+\infty$ 时, 定理 4.4.1、4.4.2 也可由定理 2.5.2、2.5.3 和 Moreau-Rockafellar 定理 4.2.3 得到. 例如, 由定理 2.5.3 可得

$$0\in\partial\Big(f+\sum_{i=1}^{p}\hat{\lambda}_i g_i+\sum_{j=1}^{q}\hat{\mu}_j h_j\Big)(\hat{x}).$$

再由 Moreau-Rockafellar 定理 4.2.3 即得(4.4.3).

由于定理 4.4.2 的重要性, 我们再在 X 为分离局部凸空间时, 给出它的两个直接证明. 一个利用 Moreau-Rockafellar 定理 4.2.3, 另一个利用 Дубовицкий-Милютин 定理 4.3.4. 由此也可看出这两条定理的作用. 两个证明都是只证明必要性.

用 Moreau-Rockafellar 定理 4.2.3 证明定理 4.4.2　令

$$G_i=\{x\in X\mid g_i(x)\leqslant 0\}\quad (i=1,\cdots,p),$$

$$G=\bigcap_{i=1}^{p}G_i=\{x\in X\mid \max_{1\leqslant i\leqslant p}g_i(x)\leqslant 0\}.$$

那么问题$(\mathscr{P})$有解$\hat{x}$等价于

$$\begin{cases}F_1(x)=f(x)-f(\hat{x})+\delta_G(x)\geqslant 0;\\ \forall x:\ h(x)=(h_1(x),\ \cdots,\ h_q(x))=0.\end{cases}$$

同定理 4.4.1 的证明一样，它等价于存在$\alpha>0$，$\hat{\mu}_1,\ \cdots,\ \hat{\mu}_q\in\mathbf{R}$，使得

$$0\in\alpha\partial F_1(\hat{x})+\sum_{j=1}^{q}\hat{\mu}_j\partial h_j(\hat{x}),$$

这里不妨取$\alpha=1$．另一方面，$\hat{x}\in\operatorname{dom}f\cap\operatorname{dom}\delta_G$，且$f$在$\hat{x}$上连续，故由 Moreau-Rockafellar 定理，有

$$\partial F_1(\hat{x})=\partial f(\hat{x})+\partial\delta_G(\hat{x}),$$

从而

$$0\in\partial f(\hat{x})+\partial\delta_G(\hat{x})+\sum_{j=1}^{q}\hat{\mu}_j\partial h_j(\hat{x}).\tag{4.4.4}$$

由条件(S)和定理 4.3.5，有

$$\partial\delta_G(\hat{x})=\begin{cases}\bigcup_{\lambda\geqslant 0}\lambda\partial\max_{1\leqslant i\leqslant p}g_i(\hat{x}),\ \text{当}\max_{1\leqslant i\leqslant p}g_i(\hat{x})=0;\\ \{0\},\qquad\qquad \text{当}\max_{1\leqslant i\leqslant p}g_i(\hat{x})<0.\end{cases}\tag{4.4.5}$$

再由定理 4.2.2，有

$$\partial\max_{1\leqslant i\leqslant p}g_i(\hat{x})=co\bigcup_{i:g_i(\hat{x})=0}\partial g_i(\hat{x}).\tag{4.4.6}$$

从而由(4.4.4)～(4.4.6)，又存在$\hat{\lambda}_1,\ \cdots,\ \hat{\lambda}_p\geqslant 0$，使得(4.4.3)、(4.4.2)成立．

注　这个证明的好处之一，在于没有条件(S)时，我们得到了一个新的充要条件 (4.4.4)；好处之二，在于我们没有利用条件(S)中的$h_j(x_0)=0$，即条件(S)实际上可以减弱．∎

用 Дубовицкий-Милютин 定理 4.3.4 证明定理 4.4.2

为简单起见，要求条件(S)中的x_0为$g_1,\ \cdots,\ g_p$的连续点，不妨假定$\hat{x}=0$．且$0\notin\partial f(0)$（否则，定理得证）．令

$$F=\{x\in X\mid f(x)<f(0)\},\ \ \tilde{F}=\bigcup_{\lambda>0}\lambda F,$$

$$G_i=\{x\in X\,|\,g_i(x)\leqslant 0\},\ \widetilde{G}_i=\bigcup_{\lambda>0}\lambda G_i\ \ (i=1,\ \cdots,\ p),$$
$$H=\{x\in X\,|\,h_j(x)=0,\ j=1,\ \cdots,\ q\}.$$

那么，$\widetilde{F}$, $\widetilde{G}_1$, $\cdots$, $\widetilde{G}_p$, H 都是凸锥，且由于 f 在 0 上连续，而 0 又不是 f 的最小值点，故存在 0 的邻域 $U\subset\operatorname{int}\operatorname{dom}f$ 以及 $x\in U$，使得 $f(x)<f(0)$. 但由定理 3.5.1, f 将也在 x 上连续，由此可见，

$$\phi\neq\operatorname{int}F\subset\operatorname{int}\widetilde{F}.$$

另一方面，条件 (S) 又说明

$$x_0\in\operatorname{ri}H\cap\left(\bigcap_{i=1}^{p}\operatorname{int}G_i\right)\subset\operatorname{ri}H\cap\left(\bigcap_{i=1}^{p}\operatorname{int}\widetilde{G}_i\right),$$

即

$$\operatorname{ri}H\cap\left(\bigcap_{i=1}^{p}\operatorname{int}\widetilde{G}_i\right)\neq\varnothing.$$

但 $\widetilde{F}\cap H\cap\left(\bigcap_{i=1}^{p}\widetilde{G}_i\right)=\varnothing$. 否则，将存在 $\tilde{x}\in X$, λ_0, λ_1, $\cdots$, $\lambda_p>0$, 满足

$$f(\tilde{x}/\lambda_0)<f(0),\ \ g_i(\tilde{x}/\lambda_i)\leqslant 0\quad (i=1,\ \cdots,\ p),$$
$$h_j(\tilde{x})=0\quad (j=1,\ \cdots,\ q).$$

令 $\lambda=\max\limits_{0\leqslant i\leqslant p}\lambda_i$, 则有

$$f(\tilde{x}/\lambda)=f\left(\frac{\lambda_0}{\lambda}\cdot\frac{\tilde{x}}{\lambda_0}\right)\leqslant\left(1-\frac{\lambda_0}{\lambda}\right)f(0)+\frac{\lambda_0}{\lambda}f(\tilde{x}/\lambda_0)<f(0),$$

$$g_i(\tilde{x}/\lambda)=g_i\left(\frac{\lambda_i}{\lambda}\cdot\frac{\tilde{x}}{\lambda_i}\right)\leqslant\left(1-\frac{\lambda_i}{\lambda}\right)g_i(0)+\frac{\lambda_i}{\lambda}g_i(\tilde{x}/\lambda_i)\leqslant 0\quad (i=1,\ \cdots,\ p),$$

$$h_j(\tilde{x}/\lambda)=\frac{1}{\lambda}h_j(\hat{x})+\left(1-\frac{1}{\lambda}\right)h_j(0)=0\quad (j=1,\ \cdots,\ q),$$

与 $\hat{x}=0$ 为解相矛盾. 这样由定理 4.3.4 注可得，存在 $x^*_{\widetilde{F}}\in\widetilde{F}^-$, $x^*_{\widetilde{F}}\neq 0$, $x^*_{\widetilde{G}_i}\in\widetilde{G}_i^-$ $(i=1,\ \cdots,\ p)$, $x^*_H\in H^-$, 使得

$$x^*_{\widetilde{F}}+\sum_{i=1}^{p}x^*_{\widetilde{G}_i}+x^*_H=0. \tag{4.4.7}$$

此外，由定理 4.3.5 易证，

$$\widetilde{F}^{-}=\bigcup_{\lambda\geq 0}\lambda\partial f(0),$$

$$\widetilde{G}_i^{-}=\begin{cases}\bigcup_{\lambda\geq 0}\lambda\partial g_i(0), & \text{当 } g_i(0)=0\\ \{0\}, & \text{当 } g_i(0)<0\end{cases}\quad (i=1,\ \cdots,\ p).$$

$$\widetilde{H}^{-}=\bigcup_{\substack{\mu_j\in\mathbf{R}\\ 1\leq j\leq q}}\sum_{j=1}^{q}\mu_j\partial h_j(0).$$

这样由(4.4.7)可得(4.4.3)、(4.4.2). ▌

注　这一证明在这里意义不大，但类似的证明也适用于非凸的光滑函数的局部极值问题，那时就能显示出这一凸锥分离定理的作用. ▌

下面讨论凸规划问题($\mathscr{P}$)的摄动问题，即当约束条件有变化时对($\mathscr{P}$)的影响问题. 设 $a=(a^i)\in\mathbf{R}^p$, $b=(b^j)\in\mathbf{R}^q$. 我们考虑下列问题:

$$(\mathscr{P}_{a,b})\begin{cases}f(x)\to\min;\\ g_i(x)+a^i\leq 0 & (i=1,\ \cdots,\ p);\\ h_j(x)+b^j=0 & (j=1,\ \cdots,\ q).\end{cases}$$

这时，问题($\mathscr{P}_{a,b}$)的值 $V=V(a,\ b)$可以看作 $\mathbf{R}^p\times\mathbf{R}^q$ 上的$(a,\ b)$的函数.

命题 4.4.3　设 f, g_1, $\cdots$, g_p 为线性空间 X 上的凸函数, h_1, $\cdots$, h_q 为 X 上的仿射函数. $V=V(a,\ b)$为问题($\mathscr{P}_{a,b}$)的值，那么 V 是 $\mathbf{R}^p\times\mathbf{R}^q$ 上的凸函数.

证明　设 $\mathrm{dom}V=\{(a,\ b)\in\mathbf{R}^p\times\mathbf{R}^q\,|\,V(a,\ b)<+\infty\}$. 在此只需证明 V 在 $\mathrm{dom}V$ 中满足凸性不等式. 设$(a_1,\ b_1)$, $(a_2,\ b_2)\in\mathrm{dom}V$, 且 $V(a_1,\ b_1)$, $V(a_2,\ b_2)\neq-\infty$. 那么，对于任何 $\varepsilon>0$, 存在 x_1、$x_2\in X$, 使得

$$\begin{aligned}&f(x_1)\leq V(a_1,\ b_1)+\varepsilon;\\ &g_i(x_1)\leq -a_1^i\quad (i=1,\ \cdots,\ p);\\ &h_j(x_1)=-b_1^j\quad (j=1,\ \cdots,\ q);\\ &f(x_2)\leq V(a_2,\ b_2)+\varepsilon;\end{aligned}$$

$$g_i(x_2) \leqslant -a_2^i \quad (i=1, \cdots, p);$$
$$h_j(x_2) = -b_2^j \quad (j=1, \cdots, q).$$

由于f、g_i等是凸函数,h_j等是仿射函数,故对于任何$\lambda \in (0, 1)$,有

$$f((1-\lambda)x_1+\lambda x_2) \leqslant (1-\lambda)f(x_1)+\lambda f(x_2)$$
$$\leqslant (1-\lambda)V(a_1, b_1)+\lambda V(a_2, b_2)+\varepsilon;$$
$$g_i((1-\lambda)x_1+\lambda x_2) \leqslant (1-\lambda)g_i(x_1)+\lambda g_i(x_2)$$
$$\leqslant -(1-\lambda)a_1^i-\lambda a_2^i \quad (i=1, \cdots, p);$$
$$h_j((1-\lambda)x_1+\lambda x_2) = (1-\lambda)h_j(x_1)+\lambda h_j(x_2)$$
$$= -(1-\lambda)b_1^j-\lambda b_2^j \quad (j=1, \cdots, q).$$

因此,$(1-\lambda)x_1+\lambda x_2$满足约束条件,从而

$$V((1-\lambda)a_1+\lambda a_2, (1-\lambda)b_1+\lambda b_2) \leqslant f((1-\lambda)x_1+\lambda x_2)$$
$$\leqslant (1-\lambda)V(a_1, b_1)+\lambda V(a_2, b_2)+\varepsilon.$$

由$\varepsilon>0$任意,即得V的凸性不等式. 当$V(a_1, b_1)$或$V(a_2, b_2)=-\infty$时,可用$-n$来代替上面的$V(a_1, b_1)+\varepsilon$或$V(a_2, b_2)+\varepsilon$,这里n为任意自然数,仍能得到V的凸性不等式. ▌

命题 4.4.4 设问题$(\mathscr{P}_{\hat{a},\hat{b}})$的 Lagrange 乘子存在,且

$$\{(\hat{\lambda}, \hat{\mu})\} \subset \mathbf{R}_+^{p*} \times \mathbf{R}^{q*}$$

为$(\mathscr{P}_{\hat{a},\hat{b}})$的 Lagrange 乘子的全体. 那么$V=V(a, b)$是$\mathbf{R}^p \times \mathbf{R}$上的真凸函数,它在$(\hat{a}, \hat{b})$上次可微,且

$$\partial V(\hat{a}, \hat{b}) = \{(\hat{\lambda}, \hat{\mu})\}.$$

证明 $(\hat{\lambda}, \hat{\mu})$是问题$(\mathscr{P}_{\hat{a},\hat{b}})$的 Lagrange 乘子意味着

$$V(\hat{a}, \hat{b}) = \inf_{x\in X}\{f(x)+\langle\hat{\lambda}, g(x)+\hat{a}\rangle+\langle\hat{\mu}, h(x)+\hat{b}\rangle\}$$
$$=\inf_{x\in X}\{f(x)+\langle\hat{\lambda}, g(x)\rangle+\langle\hat{\mu}, h(x)\rangle\}$$
$$+\langle\hat{\lambda}, \hat{a}\rangle+\langle\hat{\mu}, \hat{b}\rangle.$$

由于$\hat{\lambda} \in \mathbf{R}_+^{p*}$,可得

$$V(\hat{a}, \hat{b}) = \inf_{x\in X} \inf_{\substack{a\in\mathbf{R}^p, b\in\mathbf{R}^q \\ g(x)+a\leqslant 0, h(x)+b=0}} \{f(x)-\langle\hat{\lambda}, a\rangle-\langle\hat{\mu}, b\rangle\}$$
$$+\langle\hat{\lambda}, \hat{a}\rangle+\langle\hat{\mu}, \hat{b}\rangle$$

$$= \inf_{(a,b)\in \mathbf{R}^p\times \mathbf{R}^q} \inf_{\substack{x\in X \\ g(x)+a\leqslant 0, \\ h(x)+b=0}} \{f(x)-\langle\hat{\lambda}, a\rangle-\langle\hat{\mu}, b\rangle\}$$

$$+\langle\hat{\lambda}, \hat{a}\rangle+\langle\hat{\mu}, \hat{b}\rangle$$

$$= \inf_{(a,b)\in \mathbf{R}^p\times \mathbf{R}^q}\{V(a, b)-\langle\hat{\lambda}, a\rangle-\langle\hat{\mu}, b\rangle\}$$

$$+\langle\hat{\lambda}, \hat{a}\rangle+\langle\hat{\mu}, \hat{b}\rangle.$$

因此，

$$V(a, b)-V(\hat{a}, \hat{b})\geqslant\langle\hat{\lambda}, a-\hat{a}\rangle+\langle\hat{\mu}, b-\hat{b}\rangle,$$

$$\forall (a, b)\in\mathbf{R}^p\times\mathbf{R}^q.$$

从而 $V(a, b)\neq -\infty$，且 $(\hat{\lambda}, \hat{\mu})\in\partial V(\hat{a}, \hat{b})$.

反之，由上逆推也可得 $\partial V(\hat{a}, \hat{b})$ 中的每一元素都是 $(\mathscr{P}_{\hat{a},\hat{b}})$ 的 Lagrange 乘子. ▮

推论 如果问题 $(\mathscr{P}_{\hat{a},\hat{b}})$ 有唯一的 Lagrange 乘子 $(\hat{\lambda}, \hat{\mu})$，那么 V 在 $(\hat{a}, \hat{b})$ 上 Gâteaux 可导，即

$$\lim_{t\to 0}\frac{V(\hat{a}+ta, \hat{b}+tb)-V(\hat{a}, \hat{b})}{t}=\langle(\hat{\lambda}, \hat{\mu}), (a, b)\rangle$$

$$=\langle\hat{\lambda}, a\rangle+\langle\hat{\mu}, b\rangle. ▮$$

注 1 命题 4.4.4 及其推论对 §2.6 中的 Lagrange 乘子的经济解释又作了进一步说明. 例如，"最优利率" λ（更一般的是"影子价格"）实际上就是总收益关于资金总量（或某物资总量）的变化率. 这也就是说，在一定条件下，计划部门如果能够掌握总收益关于总资金（或物资）总量的变化情况，它就可以变集中决策为分散决策. 这个结论显然是有重要的实际意义的. ▮

注 2 与以前一样，命题 4.4.3、4.4.4 也可推广到无限维情形. 尤其是对于等式约束，这种推广是直截了当的. ▮

§4.5 凸规划的一般对偶理论

我们在前面讨论带等式约束和不等式约束的凸规划问题时，是通过引入 Lagrange 函数来得到 Lagrange 乘子和对偶规划问题的. 由此得到的对偶问题在形式上并不很对称. 现在要提出在形

式上更为对称的一般对偶理论. 为此,先回顾一下原来的讨论.

设 $f, g_1, \cdots, g_p$ 为 X 上的真凸函数. 对于下列问题(为简单起见,我们略去等式约束):

$$(\mathscr{P})\begin{cases} f(x) \to \min, \\ g_i(x) \leqslant 0 \quad (i=1, \cdots, p). \end{cases}$$

其 Lagrange 函数为

$$L(x, \lambda) = f(x) + \langle \lambda, g(x) \rangle = f(x) + \sum_{i=1}^{p} \lambda_i g_i(x).$$

于是$(\mathscr{P})$等价于

$$(\mathscr{P})\begin{cases} \sup\limits_{\lambda \in \mathbf{R}_+^{p*}} L(x, \lambda) \to \min, \\ x \in X. \end{cases}$$

其对偶问题为

$$(\mathscr{P}^*)\begin{cases} \inf\limits_{x \in X} L(x, \lambda) \to \max, \\ \lambda \in \mathbf{R}_+^{p*}. \end{cases}$$

为了使形式更对称,我们定义

$$\widetilde{L}(x, \lambda) = \begin{cases} L(x, \lambda), & \text{当 } \lambda \in \mathbf{R}_+^{p*}; \\ -\infty, & \text{当 } \lambda \notin \mathbf{R}_+^{p*}. \end{cases}$$

则$(\mathscr{P})$、$(\mathscr{P}^*)$将等价于

$$(\mathscr{P})\begin{cases} \sup\limits_{\lambda \in \mathbf{R}^{p*}} \widetilde{L}(x, \lambda) \to \min, \\ x \in X, \end{cases}$$

$$(\mathscr{P}^*)\begin{cases} \inf\limits_{x \in X} \widetilde{L}(x, \lambda) \to \max, \\ \lambda \in \mathbf{R}^{p*}. \end{cases}$$

由此自然会想到,更一般的对偶凸规划问题应该有下列形式:

$$(\mathscr{P})\begin{cases} \sup\limits_{y^* \in Y^*} L(x, y^*) \to \min, \\ x \in X. \end{cases}$$

$$(\mathscr{P}^*)\begin{cases} \inf\limits_{x \in X} L(x, y^*) \to \max, \\ y^* \in Y^*. \end{cases}$$

这里 Y^* 为另一个(拓扑)线性空间 Y 的对偶空间.

$$L:\ X\times Y^*\to \mathbf{R}\cup\{\pm\infty\}$$

为某个一般的"Lagrange 函数".

另一方面，$\sup\limits_{y^*\in Y^*}L(x,\ y^*)$又使人想起共轭函数，即它可以表为

$$\sup_{y^*\in Y^*}L(x,\ y^*)=\sup_{y^*\in Y^*}\{\langle y^*,\ y\rangle-(-L(x,\ y^*))\}_{y=0}.$$

这样，当 $-L(x,\ y^*)$关于 y^* 是下半连续凸函数时，令

$$F(x,\ y)=\sup_{y^*\in Y^*}\{\langle y^*,\ y\rangle+L(x,\ y^*)\},\tag{4.5.1}$$

那么就有($\mathscr{P}$)等价于

$$(\mathscr{P})\begin{cases}F(x,\ 0)\to\min,\\ x\in X.\end{cases}$$

同时，又有

$$\begin{aligned}\inf_{x\in X}L(x,\ y^*)&=-\sup_{x\in X}\{-L(x,\ y^*)\}\\&=-\sup_{x\in X}\{\langle x^*,\ x\rangle-L(x,\ y^*)\}_{x^*=0}.\end{aligned}$$

但由(4.5.1)又有

$$-L(x,\ y^*)=\sup_{y\in Y}\{\langle y^*,\ y\rangle-F(x,\ y)\}.$$

因此，

$$\begin{aligned}\inf_{x\in X}L(x,\ y^*)&=-\sup_{x\in X}\{\langle x^*,\ x\rangle+\sup_{y\in Y}\{\langle y^*,\ y\rangle\\&\qquad-F(x,\ y)\}\}_{x^*=0}\\&=-\sup_{x\in X}\sup_{y\in Y}\{\langle x^*,\ x\rangle+\langle y^*,\ y\rangle\\&\qquad-F(x,\ y)\}_{x^*=0}\\&=-F^*(x^*,\ y^*)\big|_{x^*=0}=-F^*(0,\ y^*).\end{aligned}$$

从而($\mathscr{P}^*$)等价于

$$(\mathscr{P}^*)\begin{cases}-F^*(0,\ y^*)\to\max,\\ y^*\in Y^*.\end{cases}$$

为了使讨论的结果有更多的应用，我们提出下列更一般的问题：设 X、Y 为两个分离局部凸空间，X^*、Y^* 分别是它们的拓扑对偶. $F:\ X\times Y\to\mathbf{R}\cup\{\pm\infty\}$ 为 $X\times Y$ 上的函数(不一定凸)，$F^*:\ X^*\times Y^*\to\mathbf{R}\cup\{\pm\infty\}$ 为其共轭函数，即

$$F^*(x^*, y^*)=\sup_{x\in X}\sup_{y\in Y}\{\langle x^*, x\rangle+\langle y^*, y\rangle-F(x, y)\}.$$

$A: X\to Y$ 为 X 到 Y 的线性映射，$A^*: Y^*\to X^*$ 为 A 的共轭映射，即它满足

$$\langle A^*y^*, x\rangle=\langle y^*, Ax\rangle, \forall x\in X, \forall y^*\in Y^*.$$

那么极值问题

$$(\mathscr{F})\begin{cases}F(x, Ax)\to\min,\\ x\in X.\end{cases}$$

的对偶问题是指

$$(\mathscr{F}^*)\begin{cases}-F^*(-A^*y^*, y^*)\to\max,\\ y^*\in Y^*.\end{cases}$$

而其二次对偶问题是指

$$(\mathscr{F}^{**})\begin{cases}F^{**}(x, Ax)\to\min,\\ x\in X.\end{cases}$$

当F为 $X\times Y$ 上的下半连续真凸函数时，$(\mathscr{F})$与$(\mathscr{F}^{**})$完全一致。在一般情况下，$(\mathscr{F}^{**})$也可称为$(\mathscr{F})$的闭凸扩张。

命题 4.5.1 $\inf_{x\in X}F(x, Ax)\geqslant\sup_{y^*\in Y^*}\{-F^*(-A^*y^*, y^*)\}.$

$$(4.5.2)$$

证明 事实上，由 Young 不等式，得

$$F(x, Ax)+F^*(-A^*y^*, y^*)\geqslant\langle -A^*y^*, x\rangle+\langle y^*, Ax\rangle$$
$$=\langle -y^*, Ax\rangle+\langle y^*, Ax\rangle=0.$$

因此，

$$\inf_{x\in X}\inf_{y^*\in Y^*}\{F(x, Ax)+F^*(-A^*y^*, y^*)\}$$
$$=\inf_{x\in X}F(x, Ax)-\sup_{y^*\in Y^*}\{-F^*(-A^*y^*, y^*)\}\geqslant 0,$$

即(4.5.2)成立。∎

这样一来，如果存在 $\hat{y}^*$ 满足

$$-F^*(-A^*\hat{y}^*, \hat{y}^*)=\inf_{x\in X}F(x, Ax), \tag{4.5.3}$$

那么由(4.5.2)可知，$\hat{y}^*$ 一定是对偶问题$(\mathscr{F}^*)$的解。满足(4.5.3)的 $\hat{y}^*$ 称为问题$(\mathscr{F})$的 Lagrange 乘子。联系前面的讨论可知，这

一名称是合理的. 同样,也可引入问题($\mathscr{F}$)的 Lagrange 函数. 但在引入 Lagrange 函数以前,先讨论一些其他结果.

命题 4.5.2 设 $F: X\times Y\to \mathbf{R}\cup\{\pm\infty\}$ 是真凸函数. 那么 $\hat{x}$ 是($\mathscr{F}$)的解、$\hat{y}^*$ 是($\mathscr{F}$)的 Lagrange 乘子的充要条件为

$$(-A^*\hat{y}^*,\ \hat{y}^*)\in\partial F(\hat{x},\ A\hat{x}). \tag{4.5.4}$$

如果 F 是下半连续真凸函数,那么(4.5.4)还等价于

$$(\hat{x},\ A\hat{x})\in\partial F^*(-A^*\hat{y}^*,\ \hat{y}^*). \tag{4.5.5}$$

证明 由定理 4.1.3, (4.5.4)等价于

$$\begin{aligned}&F(\hat{x},\ A\hat{x})+F^*(-A^*\hat{y}^*,\ \hat{y}^*)\\&\quad=\langle -A^*\hat{y}^*,\ \hat{x}\rangle+\langle\hat{y}^*,\ A\hat{x}\rangle=0.\end{aligned}$$

由(4.5.2),它也等价于

$$\begin{aligned}F(\hat{x},\ A\hat{x})&=-F^*(-A^*\hat{y}^*,\ \hat{y}^*)=\inf_{x\in X}F(x,\ Ax)\\&=\sup_{y^*\in Y^*}\{-F^*(-A^*y^*,\ y^*)\}.\end{aligned}$$

当 F 下半连续时,由定理 4.1.3, (4.5.4)等价于(4.5.5). ∎

(4.5.4)称为问题($\mathscr{F}$)的 Euler—Lagrange 方程. 这个名称来自变分学. 我们在以后要明确指出,它就是变分学中的同一方程对凸函数情形的推广.

定理 4.5.3 设 $F: X\times Y\to\mathbf{R}\cup\{\pm\infty\}$ 为真凸函数, $h: X\times Y\to Y$ 定义为

$$h(x,\ y)=Ax-y,\ \forall(x,\ y)\in X\times Y.$$

如果问题($\mathscr{F}$)的值 $V=\inf\limits_{x\in X}F(x,\ Ax)$ 有限,

$$0\in\operatorname{int}\{h(\operatorname{dom}F)\}, \tag{4.5.6}$$

且对于满足 $\operatorname{int}h(G)\neq\phi$ 的集合 $G\subset\operatorname{dom}F$, F 在 G 中上有界,那么($\mathscr{F}$)的 Lagrange 乘子存在.

证明 设 $V=\inf\limits_{x\in X}F(x,\ Ax)$. 由假设 $V\neq\pm\infty$. 于是有

$$\begin{cases}F(x,\ y)-V\geqslant 0,\\ \forall(x,\ y):h(x,\ y)=Ax-y=0.\end{cases}$$

根据定理 2.5.1, 条件(4.5.6)可导得存在 Y 上的线性形式 $\hat{y}^*$, 使

得

$$F(x, y)-V+\langle\hat{y}^*, Ax-y\rangle\geqslant 0, \ \forall(x, y)\in X\times Y. \tag{4.5.7}$$

由此也可得到

$$\langle -\hat{y}^*, Ax-y\rangle\leqslant F(x, y)-V, \ \forall(x, y)\in G.$$

由F在G中上有界，$h(G)$的内部非空，可得$-\hat{y}^*$在开集$\operatorname{int}h(G)$中上有界．因此，$\hat{y}^*$连续（定理3.5.1）即$\hat{y}^*\in Y^*$．最后，由(4.5.7)还可得

$$\begin{aligned}-F^*(-A^*\hat{y}^*, \hat{y}^*)&=-\sup_{(x,y)\in X\times Y}\{\langle -A^*y^*, x\rangle\\&\quad+\langle\hat{y}^*, y\rangle-F(x, y)\}\\&=\inf_{(x,y)\in X\times Y}\{F(x, y)+\langle\hat{y}^*, Ax-y\rangle\}\\&\geqslant V=\inf_{x\in X}F(x, Ax),\end{aligned}$$

从而由(4.5.2)、(4.5.3)可知，$\hat{y}^*$为问题($\mathscr{F}$)的 Lagrange 乘子．∎

注1　下列两条件是V有限的充分条件：

$$\exists x_0\in X, \ (x_0, Ax_0)\in\operatorname{dom}F; \tag{4.5.8}$$

$$\exists y_0^*\in Y^*, \ (-A^*y_0^*, y_0^*)\in\operatorname{dom}F^*. \tag{4.5.9}$$

事实上，这时由(4.5.2)有

$$F(x_0, Ax_0)\geqslant\inf_{x\in X}F(x, Ax)=V\geqslant -F^*(-A^*y_0^*, y_0^*). \ ∎$$

注2　如果F有形为(x, Ax)的有限连续点，而h是开映射，即它把开集变为开集，那么定理中的另两个条件是满足的．∎

注3　定理证明的中间部分及有关F的上有界假定是为指出$\hat{y}^*$的连续性．当Y有限维或赋以最细的局部凸拓扑时，其上所有的线性形式都连续，从而这些都变得不必要．因此，定理4.5.3的代数陈述较为简单．为了在形式上更为简洁，我们记$h=A\oplus(-1)$，即

$$(A\oplus(-1))(x, y)=Ax-y.$$

则(4.5.8)可写成

$$0\in(A\oplus(-1))\operatorname{dom}F, \tag{4.5.10}$$

类似地(4.5.9)可写成

$$0\in(1\oplus A^*)\operatorname{dom}F^*. \tag{4.5.11}$$

我们把(4.5.6)和(4.5.10)合写成一个条件，则可得到以下定理。

定理4.5.4 设 $F: X\times Y\to\mathbf{R}\cup\{+\infty\}$ 为真凸函数，这里 Y 上所有线性形式都连续．如果

$$0\in(1\oplus A^*)\operatorname{dom}F^*, \tag{4.5.11}$$

$$0\in\operatorname{int}(A\oplus(-1))\operatorname{dom}F, \tag{4.5.12}$$

那么问题($\mathscr{F}$)的 Lagrange 乘子存在．▌

注4 在 X、Y 都是有限维空间时，我们对问题($\mathscr{F}^*$)再用一次定理 4.5.4，就可得到下列更简洁的结果：

定理4.5.5 设 X、Y 都是有限维空间，$F: X\times Y\to\mathbf{R}\cup\{+\infty\}$ 为下半连续真凸函数．如果

$$0\in\operatorname{int}(A\oplus(-1))\operatorname{dom}F, \tag{4.5.12}$$

$$0\in\operatorname{int}(1\oplus A^*)\operatorname{dom}F^*, \tag{4.5.13}$$

那么问题($\mathscr{F}$)的解和 Lagrange 乘子都存在，即存在 $\hat{x}\in X$，$\hat{y}^*\in Y^*$，使得下列 Euler-Lagrange 方程成立：

$$(-A^*\hat{y}^*,\ \hat{y}^*)\in\partial F(\hat{x},\ A\hat{x}),$$

$$(\hat{x},\ A\hat{x})\in\partial F^*(-A^*\hat{y}^*,\ \hat{y}^*).$$ ▌

这个结果当 X、Y 都是自反 Banach 空间（即 X^*、Y^* 作为 Banach 空间时的拓扑对偶仍是 X、Y）以及 A 是 $X\to Y$ 的连续线性映射时仍成立，其证明自然要用到 Banach 空间上的下半连续真凸函数的性质（命题3.5.3的推论）．▌

在变分学和力学中，通过 Legendre 变换，Euler-Lagrange 方程可转化为等价的 Hamilton 方程．现在同样也可有这一转化：

定义 $H: X\times Y^*\to\mathbf{R}\cup\{\pm\infty\}$ 为

$$H(x,\ y^*)=\sup_{y\in Y}\{\langle y^*,\ y\rangle-F(x,\ y)\}. \tag{4.5.14}$$

它称为问题($\mathscr{F}$)的 Hamilton 函数．$H(x,\ y^*)$显然是 y^* 的下半连续凸函数．但是作为 x 的函数，当 F 关于 x 凸时，H 是 x 的凹函数；而当 F 关于 x 凹时，H 是 x 的凸函数．后一结论由凸函数族的上包络是凸函数而得到；前一结论则是下列命题的推论：

命题 4.5.6 设 $g(x, y)$ 是线性空间 $X\times Y$ 上的凸函数. 那么 $f(x)=\inf\limits_{y\in Y} g(x, y)$ 是线性空间 X 上的凸函数. ▌

证明留给读者作为习题(参看第三章习题 25).

由(4.5.14)可得

$$F^*(x^*, y^*)=\sup_{x\in X}\{\langle x^*, x\rangle+H(x, y^*)\}, \qquad (4.5.15)$$

且当 F 关于 y 下半连续凸时,有

$$F(x, y)=\sup_{y^*\in Y^*}\{\langle y^*, y\rangle-H(x, y^*)\}. \qquad (4.5.16)$$

定理 4.5.7 设真凸函数 F: $X\times Y\to \mathbf{R}\cup\{+\infty\}$ 在点 $(\hat{x}, A\hat{x})$ 上次可微. 那么 Euler-Lagrange 方程

$$(-A^*\hat{y}^*, \hat{y}^*)\in\partial F(\hat{x}, A\hat{x}) \qquad (4.5.4)$$

等价于下列 Hamilton 方程组:

$$\begin{cases} A\hat{x}\in\partial_{y^*}H(\hat{x}, \hat{y}^*), \\ A^*\hat{y}^*\in-\partial_x(-H(\hat{x}, \hat{y}^*)). \end{cases} \qquad (4.5.17)$$

这里 ∂_{y^*} 表示对 y^* 的次微分; ∂_x 表示对 x 的次微分

证明 事实上,(4.5.4)意味着

$$F(x, y)-F(\hat{x}, A\hat{x})\geqslant\langle -A\hat{y}^{**}, x-\hat{x}\rangle+\langle\hat{y}^*, y-A\hat{x}\rangle,$$
$$\forall (x, y)\in X\times Y. \qquad (4.5.18)$$

取 $x=\hat{x}$, 则得

$$F(\hat{x}, y)-F(\hat{x}, A\hat{x})\geqslant\langle\hat{y}^*, y-Ax\rangle, \ \forall y\in Y.$$

从而由 H 的定义(4.5.14),

$$H(\hat{x}, \hat{y}^*)=\sup_{y\in Y}\{\langle\hat{y}^*, y\rangle-F(\hat{x}, y)\}$$
$$=\langle\hat{y}^*, A\hat{x}\rangle-F(\hat{x}, A\hat{x}), \qquad (4.5.19)$$
$$H(\hat{x}, y^*)-H(\hat{x}, \hat{y}^*)\geqslant\langle y^*, A\hat{x}\rangle-F(\hat{x}, A\hat{x})$$
$$-\langle\hat{y}^*, A\hat{x}\rangle+F(\hat{x}, A\hat{x})$$
$$=\langle y^*-\hat{y}^*, A\hat{x}\rangle, \ \forall y^*\in Y^*.$$

因此,

$$A\hat{x}\in\partial_{y^*}H(\hat{x}, \hat{y}^*).$$

另一方面,(4.5.18)也可写成

$$\langle \hat{y}^*, y\rangle - F(x, y) \leqslant \langle A^*\hat{y}^*, x\rangle - F(\hat{x}, A\hat{x})$$
$$= \langle A^*\hat{y}^*, x - \hat{x}\rangle + H(\hat{x}, \hat{y}^*),$$
$$\forall (x, y) \in X \times Y.$$

这里后一步由(4.5.19)而得到. 从而

$$-H(x, \hat{y}^*) + H(\hat{x}, \hat{y}^*) \geqslant \langle -A^*\hat{y}^*, x - \hat{x}\rangle, \ \forall x \in X.$$

因此,

$$A^*\hat{y}^* \in -\partial_x(-H(\hat{x}, \hat{y}^*)).$$

反之,如果 $A\hat{x} \in \partial_{y^*}H(\hat{x}, \hat{y}^*)$, 那么由定理 4.1.3, 有

$$H(\hat{x}, \hat{y}^*) + H^*(\hat{x}, A\hat{x}) = \langle \hat{y}^*, A\hat{x}\rangle. \tag{4.5.20}$$

这里

$$H^*(\hat{x}, A\hat{x}) = \sup_{y^* \in Y^*}\{\langle y^*, A\hat{x}\rangle - H(\hat{x}, y^*)\} \leqslant F(\hat{x}, A\hat{x}). \tag{4.5.21}$$

但另一方面, F 在点$(\hat{x}, A\hat{x})$上次可微, 特别是, 作为 y 的函数, $F(\hat{x}, y)$在 $y = A\hat{x}$ 上次可微,从而存在 $\hat{y}^* \in \partial_y F(\hat{x}, A\hat{x})$, 使得

$$F(\hat{x}, A\hat{x}) + H(\hat{x}, \hat{y}^*) = \langle \hat{y}^*, A\hat{x}\rangle.$$

因此,又有

$$F(\hat{x}, A\hat{x}) = \langle \hat{y}^*, A\hat{x}\rangle - H(\hat{x}, \hat{y}^*) \leqslant H^*(\hat{x}, A\hat{x}). \tag{4.5.22}$$

这样,由(4.5.20)及(4.5.21), 我们得到 $H^*(\hat{x}, A\hat{x}) = F(\hat{x}, A\hat{x})$, 以及

$$\langle \hat{y}^*, A\hat{x}\rangle = F(\hat{x}, A\hat{x}) + H(\hat{x}, \hat{y}^*).$$

从而再由(4.5.14)可得 $F(x, y) \geqslant \langle \hat{y}^*, y\rangle - H(x, \hat{y}^*)$, 以至

$$F(x, y) - F(\hat{x}, A\hat{x}) \geqslant \langle \hat{y}^*, y - A\hat{x}\rangle - H(x, \hat{y}^*) + H(\hat{x}, \hat{y}^*),$$
$$\forall x \in X, \ \forall y \in Y.$$

另一方面, 由 $A^*\hat{y}^* \in -\partial_x(-H(\hat{x}, \hat{x}^*))$, 又可得

$$-H(x, \hat{y}^*) + H(\hat{x}, \hat{y}^*) \geqslant \langle -A^*\hat{y}^*, x - \hat{x}\rangle, \ \forall x \in X.$$

因此

$$F(x, y) - F(\hat{x}, A\hat{x}) \geqslant \langle \hat{y}^* - A\hat{x}\rangle + \langle -A^*\hat{y}^*, x - \hat{x}\rangle,$$
$$\forall x \in X, \ \forall y \in Y,$$

即(4.5.4)成立. ∎

最后，我们定义问题 $(\mathscr{F})$ 的 Lagrange 函数 L: $X\times Y^*\to\mathbf{R}\cup\{\pm\infty\}$ 为

$$L(x,\ y^*)=\langle y^*,\ Ax\rangle-H(x,\ y^*) \tag{4.5.23}$$

当 F 关于 x 凸时，它是 x 的凸函数．但它对 y^* 总是 y^* 的凹函数．

注 由于目前的理论有几个不同的来源，这里的名词与它们的原型之间的关系有点混乱．上面定义的 Lagrange 函数由下面的定理可知，它对应凸规划问题中的 Lagrange 函数（起源于约束极值问题）．但是变分学和力学中也有 Lagrange 函数的名称，在这里它相当于当 $F(x,\ Ax)=\int_0^T L(t,\ x(t),\ \dot{x}(t))dt$（$Ax=\dot{x}$ 为 x 关于 t 的导数）时的被积函数 $L(t,\ x(t),\ \dot{x}(t))$．上面定义的 Lagrange 函数当然不是指 $L(t,\ x(t),\ \dot{x}(t))$．此外，在变分学和力学中，$L(t,\ x,\ \dot{x})$ 的 Legendre 变换（关于 $\dot{x}$ 的共轭函数）也称为 Hamilton 函数．与这里定义的 Hamilton 函数也有区别．以后要进一步阐明其中的关系．■

定理 4.5.8 设 F: $X\times Y\to\mathbf{R}\cup\{\pm\infty\}$ 为任意函数．$L(x,\ y^*)$ 为对应的问题 $(\mathscr{F})$ 的 Lagrange 函数．那么

$$-F^*(-Ay^*,\ y^*)=\inf_{x\in X}L(x,\ y^*), \tag{4.5.24}$$

$$F(x,\ Ax)\geqslant\sup_{y^*\in Y^*}L(x,\ y^*). \tag{4.5.25}$$

如果 F 关于 y 是下半连续真凸函数，那么也有

$$F(x,\ Ax)=\sup_{y^*\in Y^*}L(x,\ y^*), \tag{4.5.26}$$

且 $\hat{x}$ 为 $(\mathscr{F})$ 的解、$\hat{y}^*$ 为 $(\mathscr{F})$ 的 Lagrange 乘子的充要条件为 $(\hat{x},\ \hat{y}^*)$ 是 L 的鞍点，即

$$L(\hat{x},\ y^*)\leqslant L(\hat{x},\ \hat{y}^*)\leqslant L(x,\ \hat{y}^*) \tag{4.5.27}$$

$$\forall x\in X,\ \forall y^*\in Y^*.$$

证明 事实上，由定义 (4.5.23) 和 (4.5.15)，

$$\inf_x L(x,\ y^*)=\inf_x\{\langle y^*,\ Ax\rangle-H(x,\ y^*)\}$$

$$=-\sup_x\{\langle -Ay^*,\ x\rangle+H(x,\ y^*)\}$$

$$= -F^*(-Ay^*, y^*),$$

即(4.5.24)成立．而由(4.5.14)，

$$F(x, y) \geqslant \langle y^*, y \rangle - H(x, y^*), \ \forall y^* \in Y^*.$$

从而

$$F(x, Ax) \geqslant \sup_{y^*} \{\langle y^*, Ax \rangle - H(x, y^*)\} = \sup_{y^*} L(x, y^*).$$

当F对y下半连续真凸时，由(4.5.16)，上式中的等号成立．因此，(4.5.25)与(4.5.26)成立．

如果(4.5.24)与(4.5.26)成立，那么$(\hat{x}, \hat{y}^*)$是L的鞍点等价于

$$L(\hat{x}, \hat{y}^*) = \sup_{y^* \in Y^*} \inf_{x \in X} L(x, y^*) = \inf_{x \in X} \sup_{y^* \in Y^*} L(x, y^*)$$
$$= \sup_{y^* \in Y^*} L(\hat{x}, y^*) = \inf_{x \in X} L(x, \hat{y}^*);$$

由此易证它等价于$\hat{x}$是$(\mathscr{F})$的解、$\hat{y}^*$是$(\mathscr{F}^*)$的 Lagrange 乘子．■

注 上述鞍点在定理中除假定F对y下半连续真凸外，没有任何其他假定．回顾定理 2.4.3，那里对$f(x)$、$g_i(x)$、$h_j(x)$等也没有任何假定．■

下面讨论问题$(\mathscr{F})$的摄动问题：

$$(\mathscr{F}_{x^*, y})\begin{cases} F(x, Ax+y) - \langle x^*, x \rangle \to \min, \\ x \in X. \end{cases}$$

它的对偶问题将是

$$(\mathscr{F}^*_{x^*, y})\begin{cases} -F^*(-A^*y^*+x^*, y^*) + \langle y^*, y \rangle \to \max, \\ y^* \in Y^*. \end{cases}$$

对于问题$(\mathscr{F}_{x^*, y})$和$(\mathscr{F}^*_{x^*, y})$，仍可与上述有平行的定义和结果．例如，(4.5.2)变为

$$\inf_{x \in X} \{F(x, Ax+y) - \langle x^*, x \rangle\}$$
$$\geqslant \sup_{y^* \in Y^*} \{-F^*(-A^*y^*+x^*, y^*) + \langle y^*, y \rangle\}.$$

Euler-Lagrange 方程(4.5.4)与(4.5.5)变为

$$(-A_T\hat{y}^*+x^*, \hat{y}^*) \in \partial F(\hat{x}, A\hat{x}+y),$$
$$(\hat{x}, A\hat{x}+y) \in \partial F^*(-A^*\hat{y}^*+x^*, \hat{y}^*).$$

条件(4.5.12)与(4.5.13)变为

$$-y\in\operatorname{int}(A\oplus(-1))\operatorname{dom}F,$$

$$x^*\in\operatorname{int}(1\oplus A^*)\operatorname{dom}F^*.$$

Hamilton 方程组(4.5.17)变为

$$\begin{cases}A\hat{x}+y\in\partial_{y^*}H(\hat{x},\ \hat{y}^*))\\ A^*\hat{y}^*-x^*\in-\partial_x(-H(\hat{x},\ \hat{y}^*)).\end{cases}$$

这样，可以有下列一般的结果：

定理 4.5.5′ 设X、Y 都是有限维空间(更一般的可以是自反 Banach 空间)，F: $X\times Y\to\mathbf{R}\cup\{+\infty\}$为下半连续真凸函数．如果

$$-y\in\operatorname{int}(A\oplus(-1))\operatorname{dom}F,$$

$$x^*\in\operatorname{int}(1\oplus A^*)\operatorname{dom}F^*.$$

那么问题$(\mathscr{F}_{x^*,y})$的解和 Lagrange 乘子都存在，即存在$\hat{x}\in X$，$\hat{y}^*\in Y$，使得下列 Euler-Lagrange 方程成立：

$$(-A^*\hat{y}^*+x^*,\ \hat{y}^*)\in\partial F(\hat{x},\ A\hat{x}+y),$$

$$(\hat{x},\ A\hat{x}+y)\in\partial F^*(-A^*\hat{y}^*+x^*\hat{y}^*).$$

或下列 Hamilton 方程组成立：

$$\begin{cases}A\hat{x}+y\in\partial_{y^*}H(\hat{x},\ \hat{y}^*),\\ A^*\hat{y}^*-x^*\in-\partial_x(-H(\hat{x},\ \hat{y}^*)).\end{cases}$$ ∎

引入摄动问题的主要目的不是把前面的结果一般化，而是要研究原问题的解与 Lagrange 乘子同摄动参数(x^*,y)之间的关系．令问题$(\mathscr{F}_{x^*,y})$的值函数为V_{x^*}: $Y\to\mathbf{R}\cup\{\pm\infty\}$

$$V_{x^*}(y)=\inf_{x\in X}\{F(x,\ Ax+y)-\langle x^*,\ x\rangle\}.\qquad(4.5.28)$$

命题 4.5.9 V_{x^*}的共轭函数为

$$V_{x^*}^*(y^*)=F^*(-Ay^*+x^*,\ y^*).\qquad(4.5.29)$$

证明 事实上，

$$V_{x^*}^*(y^*)=\sup_{y\in Y}\{\langle y^*,\ y\rangle-V_x^*(y)\}$$

$$=\sup_{y\in Y}\{\langle y^*,\ y\rangle-\inf_{x\in X}[F(x,\ Ax+y)-\langle x^*,\ x\rangle]\}$$

$$=\sup_{x\in X,y\in Y}\{\langle y^*,\ y\rangle+\langle x^*,\ x\rangle-F(x,\ Ax+y)\}$$

$$= \sup_{x\in X, y\in Y} \{\langle y^*, Ax+y\rangle + \langle -A^*y^*+x^*, x\rangle - F(x, Ax+y)\}$$

$$= \sup_{x\in X, y\in Y} \{\langle y^*, y\rangle + \langle -A^*y^*+x^*, x\rangle - F(x, y)\}$$

$$= F^*(-Ay^*+x^*, y^*).$$ ▮

命题 4.5.10 如果 F 是 $X\times Y$ 上的凸函数，那么 $V_{x^*}(y)$ 是 Y 上的凸函数. ▮

这是命题 4.5.6 的推论.

定理 4.5.11 设 F 为 $X\times Y$ 上的凸函数. 那么 $\hat{y}^*$ 是问题 $(\mathscr{F}_{x^*,y})$ 的 Lagrange 乘子，当且仅当

$$\hat{y}^* \in \partial V_{x^*}(y). \tag{4.5.30}$$

如果 $V_{x^*}(y) = V_{x^*}^{**}(y)$，那么 $\hat{y}^*$ 是问题 $(\mathscr{F}_{x^*,y})$ 的 Lagrange 乘子当且仅当 $\hat{y}^*$ 是对偶问题 $(\mathscr{F}_{x^*,y}^*)$ 的解.

证明 事实上，由定理 4.1.3，(4.5.30) 等价于

$$V_{x^*}(y) + V_{x^*}^*(\hat{y}^*) = \langle \hat{y}^*, y\rangle,$$

由命题 4.5.9，上式等价于

$$-F^*(-A^*\hat{y}^*+x^*, \hat{y}^*) + \langle \hat{y}^*, y\rangle = V_{x^*}(y)$$

$$= \inf_{x\in X} \{F(x, Ax+y) - \langle x^*, x\rangle\},$$

即 $\hat{y}^*$ 是 $(\mathscr{F}_{x^*,y})$ 的 Lagrange 乘子.

如果 $V_{x^*}(y) = V_{x^*}^{**}(y)$，那么 (4.5.30) 还等价于

$$y \in \partial V_{x^*}^*(\hat{y}^*),$$

或

$$0 \in \partial V_{x^*}^*(\hat{y}^*) - y.$$

而它等价于

$$V_{x^*}^*(\hat{y}^*) - \langle \hat{y}^*, y\rangle = \inf_{y^*\in Y^*} \{V_{x^*}^*(y^*) - \langle y^*, y\rangle\},$$

即

$$-F^*(-A\hat{y}^*+x^*, \hat{y}^*) + \langle \hat{y}^*, y\rangle$$

$$= \sup_{y^*\in Y^*} \{-F(-Ay^*+x^*, y^*) + \langle y^*, y\rangle\}.$$ ▮

当 F 为下半连续真凸函数时，可以得到定理 4.5.11 的对偶结

果如下.

定理 4.5.12 设 F 为 $X\times Y$ 上的下半连续真凸函数,

$$U_y^*(x^*)=\inf_{y^*\in Y^*}\{F^*(-A^*y^*+x^*,\ y^*)-\langle y^*,\ y\rangle\}. \tag{4.5.31}$$

那么 $\hat{x}$ 是问题 $(\mathscr{F}_{x^*,y}^*)$ 的 Lagrange 乘子,当且仅当

$$\hat{x}\in\partial U_y^*(x^*).$$

如果 U_y^* 在 x^* 上次可微,那么 $\partial U_{y*}^*(x^*)$ 就是问题 $(\mathscr{F}_{x^*,y})$ 的解的全体. ▮

这里只需注意到 $(\mathscr{F}_{x^*,y}^{**})$ 与 $(\mathscr{F}_{x^*,y})$ 实际上是一回事. 同时,由定理 4.1.3 的推论 2, U_y^* 在 x^* 上次可微是 $U_y^*(x^*)=U_y^{***}(x^*)$ 的充分条件.

下面还要指出有关 Lagrange 函数的结果.

命题 4.5.13 如果 F 关于 y 下半连续凸,那么

$$F(x,\ Ax+y)=\sup_{y^*\in Y^*}\{\langle y^*,\ y\rangle+L(x,\ y^*)\}; \tag{4.5.32}$$

$$-L(x,\ y^*)=\sup_{y\in Y}\{\langle y^*,\ y\rangle-F(x,\ Ax+y)\}; \tag{4.5.33}$$

$$V_{x^*}(y)=\inf_{x\in X}\sup_{y^*\in Y^*}\{\langle y^*,\ y\rangle-\langle x^*,\ x\rangle+L(x,\ y^*)\}; \tag{4.5.34}$$

$$U_y^*(x^*)=\inf_{y^*\in Y^*}\sup_{x\in X}\{\langle x^*,\ x\rangle-\langle y^*,\ y\rangle-L(x,\ y^*)\}; \tag{4.5.35}$$

$$V_{x^*}^*(y^*)=\sup_{x\in X}\{\langle x^*,\ x\rangle-L(x,\ y^*)\}=F^*(-A^*y^*+x^*,\ y^*); \tag{4.5.36}$$

$$U_y^{**}(x)=\sup_{y^*\in Y^*}\{\langle y^*,\ y\rangle+L(x,\ y^*)\}=F(x,\ Ax+y). \tag{4.5.37}$$

证明 由于 F 关于 y 下半连续凸,故由(4.5.16)和(4.5.23),

$$F(x,\ Ax+y)=\sup_{y^*\in Y^*}\{\langle y^*,\ y\rangle+\langle y^*,\ Ax\rangle-H(x,\ y^*)\}=\sup_{y^*\in Y^*}\{\langle y^*,\ y\rangle+L(x,\ y^*)\}.$$

即(4.5.32)成立. (4.5.33)则由(4.5.32)可得, 因为 $-L(x, y^*)$ 关于 y^* 下半连续凸. (4.5.34)则可由(4.5.32)和(4.5.28)可得. 从而

$$\begin{aligned} V_{x^*}^*(y^*) &= \sup_{y\in Y}\{\langle y^*, y\rangle - V_{x^*}(y)\} \\ &= \sup_{y\in Y}\{\langle y^*, y\rangle - \inf_{x\in X}[F(x, Ax+y) - \langle x^*, x\rangle]\} \\ &= \sup_{x\in X, y\in Y}\{\langle x^*, x\rangle + \langle y^*, y\rangle - F(x, Ax+y)\}. \end{aligned}$$

再由(4.5.33)和(4.5.29)即得(4.5.36). (4.5.35)和(4.5.37)的推导是类似的. ∎

§4.6 一般对偶理论的应用

在这一节中用例子来说明上节中得到的结果.

1° 通常的凸规划

再来讨论以前的凸规划, 即问题($\mathscr{P}$):

$$(\mathscr{P})\begin{cases} f(x)\to \min; \\ g_i(x)\leqslant 0 \quad (i=1, \cdots, p); \\ h_j(x)=0 \quad (j=1, \cdots, q). \end{cases}$$

取 $Y=\mathbf{R}^p\times\mathbf{R}^q$, $Y^*=\mathbf{R}^{p*}\times\mathbf{R}^{q*}$, $y=(a, b)$, 其中 $a\in\mathbf{R}^p$, $b\in\mathbf{R}^q$, $y^*=(\lambda, \mu)$, 其中 $\lambda\in\mathbf{R}^{p*}$, $\mu\in\mathbf{R}^{q*}$. 已知其 Lagrange 函数为

$$L(x, y^*)=L(x, \lambda, \mu)=\begin{cases} f(x)+\langle\lambda, g(x)\rangle+\langle\mu, h(x)\rangle, \\ \quad 当\ \lambda\in\mathbf{R}_+^{p*}; \\ -\infty, \quad 当\ \lambda\notin\mathbf{R}_+^{p*}. \end{cases}$$

或

$$L(x, \lambda, \mu)=f(x)+\langle\lambda, g(x)\rangle+\langle\mu, h(x)\rangle-\delta_{C^*}(\lambda),$$

这里 $C^*=\mathbf{R}_+^{p*}$. 由(4.5.32), 且令 $A=0$, 则有

$$\begin{aligned} F(x, y)=F(x, a, b) &= \sup_{(\lambda,\mu)\in Y^*}\{\langle\lambda, a\rangle+\langle\mu, b\rangle+ \\ &\qquad L(x, \lambda, \mu)\} \\ &= \sup_{\lambda\in\mathbf{R}^{p*},\mu\in\mathbf{R}^{q*}}\{f(x)+\langle\lambda, g(x)+a\rangle \end{aligned}$$

$$+\langle\mu,\ h(x)+b\rangle-\delta_{C^*}(x)\}$$
$$=f(x)+\delta_C(g(x)+a)+\delta_0(h(x)+b),$$

这里 $C=-\mathbf{R}_+^p$. 因此,$(\mathscr{P})$的一般摄动问题应为

$(\mathscr{P}_{x^*,a,b})$

$$f(x)+\delta_C(g(x)+a)+\delta_0(h(x)+b)-\langle x^*,\ x\rangle\to\min$$

或

$(\mathscr{P}_{x^*,a,b})$

$$\begin{cases}f(x)-\langle x^*,\ x\rangle\to\min;\\ g(x)+a\in-\mathbf{R}_+^p \text{ 或 } g_i(x)+a^i\leqslant 0\ \ (i=1,\ \cdots,\ p);\\ h(x)+b=0 \text{ 或 } h_j(x)+b^j=0\ \ (j=1,\ \cdots,\ q).\end{cases}$$

由(4.5.36),得

$$F^*(x^*,\ y^*)=F^*(x^*,\ \lambda,\ \mu)=\sup_{x\in X}\{\langle x^*,\ x\rangle-L(x,\ y^*)\}$$
$$=\sup_{x\in X}\{\langle x^*,\ x\rangle-f(x)-\langle\lambda,\ g(x)\rangle$$
$$-\langle\mu,\ h(x)\rangle+\delta_{C^*}(\lambda)\}$$
$$=-\inf_{x\in X}\{f(x)-\langle x^*,\ x\rangle+\langle\lambda,\ g(x)\rangle$$
$$+\langle\mu,\ h(x)\rangle-\delta_{C^*}(\lambda)\}.$$

故对偶问题应为

$(\mathscr{P}^*_{x^*,a,b})$

$$\inf_{x\in X}\{f(x)-\langle x^*,\ x\rangle+\langle\lambda,\ g(x)+a\rangle$$
$$+\langle\mu,\ h(x)+b\rangle-\delta_{C^*}(\lambda)\}\to\max,$$

$(\mathscr{P}^*_{x^*,a,b})$

$$\begin{cases}\inf_{x\in X}\{f(x)-\langle x^*,\ x\rangle+\langle\lambda,\ g(x)+a\rangle\\ \quad+\langle\mu,\ h(x)+b\rangle\}\to\max;\\ \lambda\in\mathbf{R}_+^{p*},\ \mu\in\mathbf{R}^{q*}.\end{cases}$$

现在对这些问题应用上节的结果. 由于 $Y=\mathbf{R}^p\times\mathbf{R}^q$ 为有限维空间,在对问题$(\mathscr{P})$应用定理 4.5.5 或 4.5.6 时,可以不必验证 F 的上有界条件.

定理 4.6.1 设 $f,\ g_1,\ \cdots,\ g_p$ 为 X 上的真凸函数,$h_1,\ \cdots,\ h_q$

为 X 上的仿射函数. 如果问题 $(\mathscr{P})$ 的值 $V\neq\pm\infty$, 且存在 $x_0\in\operatorname{dom}f$, 使得 f, g_1, $\cdots$, g_p; h_1, $\cdots$, h_q 都在 x_0 上连续, 同时满足条件

$$(S)\quad g_i(x_0)<0\ (i=1,\ \cdots,\ p);\ h_j(x_0)=0\ (j=1,\ \cdots,\ q),$$

则问题$(\mathscr{P})$的 Lagrange 乘子存在.

证明 只需验证对问题$(\mathscr{P})$(4.5.6)或(4.5.12)成立. 而注意到对应的 F 的表达式以及 $A=0$, 它们即

$$0\in\operatorname{int}\{a\in\mathbf{R}^p\,|\,\exists x\in\operatorname{dom}f,\ g(x)+a\leqslant 0\},\qquad(4.6.1)$$

$$0\in\operatorname{int}\{b\in\mathbf{R}^q\,|\,\exists x\in\operatorname{dom}f,\ h(x)+b=0\}.\qquad(4.6.2)$$

定理中的条件恰好能保证这两个条件当 h 是 $X\to\mathbf{R}^q$ 的满射时成立; 当 h 不是满射时, 用 $h(X)(\neq\varnothing)$代替 $\mathbf{R}^q$, 仍能得到同样的结论. ∎

这一定理可以与定理 2.5.3 和定理 4.4.2 进行比较. 它们各有各的特点. 不过定理 4.6.1 实质上还是相当于定理 2.5.3 的注 3 中所提到的情形.

对于问题$(\mathscr{P})$的 Euler-Lagrange 方程为

$$(0,\ \hat{\lambda},\ \hat{\mu})\in\partial F(\hat{x},\ 0,\ 0).$$

它等价于

$$F(\hat{x},\ 0,\ 0)+F^*(0,\ \hat{\lambda},\ \hat{\mu})=0,$$

由 F 与 F^* 的表达式可得到此式, 即

$$f(\hat{x})-\inf_{x\in X}\{f(x)+\langle\hat{\lambda},\ g(x)\rangle+\langle\hat{\mu},\ h(x)\rangle\}=0.$$

它恰好就是原先的 Lagrange 乘子的定义, 并蕴含松紧关系

$$\langle\hat{\lambda},\ g(\hat{x})\rangle=0.$$

另一方面, 又有 $F(\hat{x},\ 0,\ 0)=f(\hat{x})=V(0)$, $F^*(0,\ \hat{\lambda},\ \hat{\mu})=V^*(\hat{\lambda},\ \hat{\mu})$(参看(4.5.29)). 故目前的 Euler-Lagrange 方程即

$$V(0)+V^*(\hat{\lambda},\ \hat{\mu})=0,$$

它等价于

$$\{(\hat{\lambda},\ \hat{\mu})\}=\partial V(0).$$

这也可由定理 4.5.13 导出, 而它恰好就是命题 4.4.4 的特例.

对于问题$(\mathscr{P})$的 Hamilton 方程组为

$$0\in\partial_{(\lambda,\mu)}H(\hat{x},\ \hat{\lambda},\ \hat{\mu}),$$
$$0\in\partial_x(-H(\hat{x},\ \hat{\lambda},\ \hat{\mu})),$$

即

$$H(x,\ \hat{\lambda},\ \hat{\mu})\leqslant H(\hat{x},\ \hat{\lambda},\ \hat{\mu})\leqslant H(\hat{x},\ \lambda,\ \mu)$$
$$\forall x\in X,\ \forall\lambda\in\mathbf{R}^{p*},\ \forall\mu\in\mathbf{R}^{q*}.$$

由 $H(x,\ \lambda,\ \mu)=-L(x,\ \lambda,\ \mu)$，它也就是鞍点关系

$$L(\hat{x},\ \lambda,\ \mu)\leqslant L(\hat{x},\ \hat{\lambda},\ \hat{\mu})\leqslant L(x,\ \hat{\lambda},\ \hat{\mu}),$$
$$\forall x\in X,\ \forall\lambda\in\mathbf{R}^{p},\ \forall\mu\in\mathbf{R}^{q}.$$

同样，对于一般的问题$(\mathscr{P}_{x^*,a,b})$的 Euler–Lagrange 方程可导出

$$f(\hat{x})-\langle x^*,\ \hat{x}\rangle-\inf_{x\in X}\{f(x)-\langle x^*,\ x\rangle+\langle\hat{\lambda},\ g(x)+a\rangle$$
$$+\langle\hat{\mu},\ h(x)+b\rangle\}=0,$$
$$\langle\hat{\lambda},\ g(\hat{x})+a\rangle=0,$$
$$\{(\hat{\lambda},\ \hat{\mu})\}=\partial V_{x^*}(a,\ b).$$

而 Hamilton 方程组

$$(a,\ b)\in\partial_{(\lambda,\mu)}H(\hat{x},\ \hat{\lambda},\ \hat{\mu}),$$
$$x^*\in\partial_x(-H(\hat{x},\ \hat{\lambda},\ \hat{\mu})).$$

其鞍点定理的形式将由(4.5.34)与(4.5.35)给出。

2° 线性规划

令 $X=\mathbf{R}^n$, $Y=\mathbf{R}^p$, A: $\mathbf{R}^n\to\mathbf{R}^p$ 为 $n\times p$ 矩阵(a_i^j), f: $\mathbf{R}^n\to\mathbf{R}\cup\{+\infty\}$ 为

$$f(x)=\delta_{\mathbf{R}^n_+}(x)=\begin{cases}0, & \text{当 } x\geqslant 0\Leftrightarrow x^i\geqslant 0\ (i=1,\ \cdots,\ n);\\ +\infty, & \text{其他}.\end{cases}$$

g: $\mathbf{R}^p\to\mathbf{R}\cup\{+\infty\}$ 为

$$g(y)=\delta_{\mathbf{R}^p_+}(y)=\begin{cases}0, & \text{当 } y\geqslant 0\Leftrightarrow y^j\geqslant 0\ (j=1,\ \cdots,\ p);\\ +\infty, & \text{其他}.\end{cases}$$

则

$$f^*(x^*)=\sup_{x\in\mathbf{R}^n_+}\langle x^*,\ x\rangle=\begin{cases}0, & \text{当 } x^*\leqslant 0;\\ +\infty, & \text{其他}.\end{cases}$$

即 $f^*(x^*)=\delta_{-\overline{\mathbf{R}}^n_+}(x^*)$,

$$g^*(y^*)=\sup_{y\in\mathbf{R}^p_+}\langle y^*,y\rangle=\delta_{-\mathbf{R}^p_+}(y^*)=\begin{cases}0, & \text{当 } y^*\leqslant 0;\\ +\infty, & \text{其他}.\end{cases}$$

最后取 $F(x,y)=f(x)+g(y)$, 则 $F^*(x^*,y^*)=f^*(x^*)+g^*(y^*)$. 这时问题 $(\mathscr{F}_{x^*,y})$ 和 $(\mathscr{F}^*_{x^*,y})$ 就变为

$(\mathscr{L}_{x^*,y})$

$$f(x)+g(Ax+y)-\langle x^*,x\rangle\to\min,$$

或

$$(\mathscr{L}_{x^*,y})\begin{cases}-\langle x^*,x\rangle_n\to\min;\\ x\geqslant 0;\\ Ax\geqslant -y.\end{cases}$$

以及

$(\mathscr{L}^*_{x^*,y})$

$$-f^*(-A^*y^*+x^*)-g^*(y^*)+\langle y^*,y\rangle\to\max$$

或

$(\mathscr{L}^*_{x^*,y})$

$$\begin{cases}\langle y^*,y\rangle_p\to\max;\\ y^*\leqslant 0;\\ A^*y^*\geqslant x^*.\end{cases}$$

$(\mathscr{L}_{x^*,y})$ 与 $(\mathscr{L}^*_{x^*,y})$ 是一对互为对偶的线性规划. 取 $b=-x^*$, $c=-y$, $\lambda=-y^*$, 它们就归结为 § 2.6 的 $(\mathscr{L})$ 和 $(\mathscr{L}^*)$.

利用定理 4.5.7′, 可以得到下列结果:

命题 4.6.2 如果下列条件满足:

$$\exists x\in\mathbf{R}^n,\ x\geqslant 0,\ Ax+y>0; \tag{4.6.3}$$

$$\exists y^*\in\mathbf{R}^{p*},\ y^*\leqslant 0,\ A^*y^*-x^*>0, \tag{4.6.4}$$

那么问题 $(\mathscr{L}_{x^*,y})$ 和 $(\mathscr{L}^*_{x^*,y})$ 的解都存在, 且它们互为对方的 Lagrange 乘子. ▮

证明只需注意到 (4.6.3) 和 (4.6.4) 是定理 4.5.7′ 中的充分条件. 这个结果并不比定理 2.6.1 强, 相反, 对于定理 2.6.1, 只需下列两条件成立:

$$\exists x\in\mathbf{R}^n,\ x\geqslant 0,\ Ax+y\geqslant 0,$$
$$\exists y^*\in\mathbf{R}^{p*},\ y^*\leqslant 0,\ A^*y^*-x^*\geqslant 0.$$

这两个条件相当于定理 2.6.1 中的 c)，而只利用定理 4.5.7′ 是得不出这两个条件的．其原因在于这里始终没有考虑定理 2.5.3 后注 5 中所提到的情况．这也说明前面展开的讨论对于有仿射不等式约束的规划问题可以有一定的改进．

对于$(\mathscr{L}^{k}_{x^*,y})$的 Euler-Lagrange 方程或 Hamilton 方程组可导得

$$-A^*\hat{y}^*+x^*\in\partial f(\hat{x}),\quad \hat{x}\in\partial f^*(-A^*\hat{y}^*+x^*),\qquad(4.6.5)$$
$$A\hat{x}+y\in\partial g^*(\hat{y}^*),\qquad \hat{y}^*\in\partial g(A\hat{x}+y).\qquad(4.6.6)$$

由于

$$\partial f(\hat{x})=\partial\delta_{\mathbf{R}^n_+}(\hat{x})=\{x^*\in\mathbf{R}^{n*}\mid\langle x^*,\ x-\hat{x}\rangle\leqslant 0,\ \forall x\in\mathbf{R}^n_+\},$$
$$\partial g^*(\hat{y}^*)=\partial\delta_{\mathbf{R}^{p*}_-}(\hat{y}^*)=\{y\in\mathbf{R}^p\mid\langle y^*-\hat{y}^*,\ y\rangle\leqslant 0,$$
$$\forall y^*\in-\mathbf{R}^{p*}_+\}.$$

故由(4.6.5)可得

$$\langle -A^*\hat{y}^*+x^*,\ x-\hat{x}\rangle\leqslant 0,\ \forall x\in\mathbf{R}^n_+.\qquad(4.6.7)$$

但由于 0、$2\hat{x}\in\mathbf{R}^n_+$，故令 $x=0$ 和 $2\hat{x}$，由(4.6.7)可得

$$\pm\langle -A^*\hat{y}^*+x^*,\ \hat{x}\rangle\leqslant 0,\quad 即\quad \langle -A^*\hat{y}^*+x^*,\ \hat{x}\rangle=0.$$

从而

$$\langle -A^*\hat{y}^*+x^*,\ x\rangle\leqslant 0,\ \forall x\in\mathbf{R}^n_+,\quad 即\quad -A^*\hat{y}^*+x^*\in-\mathbf{R}^{n*}_+.$$

因此(4.6.7)就是松紧关系：

$$\langle -A^*\hat{y}^*+x^*,\ \hat{x}\rangle=0.$$

同样，(4.6.6)也就是松紧关系

$$\langle\hat{y}^*,\ A\hat{x}+y\rangle=0.$$

不过(4.6.5)、(4.6.6)还能说明解、Lagrange 乘子与摄动问题的值函数之间的关系，即

$$\hat{x}\in\partial U^*_y(x^*)=\partial f^*(-A^*\hat{y}^*+x^*);$$
$$\hat{y}^*\in\partial V_{x^*}(y)=\partial g(A\hat{x}+y).$$

3° Fenchel 问题

当 $F(x,\ y)$形式为 $F(x,\ y)=f(x)+g(y)$时，问题$(\mathscr{F})$就称

为 **Fenchel** 问题. 上面的讨论指出了线性规划问题无非是 Fenchel 问题的特例. 利用上一节的讨论, 可以得到下列一般的定理.

定理 4.6.3(Fenchel-Aubin) 设 X, Y 都是有限维空间. $f: X \to \mathbf{R}\cup\{+\infty\}$, $g: Y \to \mathbf{R}\cup\{+\infty\}$ 分别是 X、Y 上的下半连续真凸函数. $A: X \to Y$ 为线性映射. 考虑下列一般的 Fenchel 问题:

$(\mathscr{H}_{x^*,y})$

$$f(x)+g(Ax+y)-\langle x^*, x\rangle \to \min, \ x\in X,$$

$(\mathscr{H}^*_{x^*,y})$

$$-f^*(-A^*y^*+x^*)-g^*(y^*)+\langle y^*, y\rangle \to \max, \ y^*\in Y^*.$$

则有下列结果:

i) 如果

$$y\in \operatorname{int}(\operatorname{dom} g-A\operatorname{dom} f), \tag{4.6.8}$$

$$x^*\in \operatorname{dom} f^*+A^*\operatorname{dom} g^*. \tag{4.6.9}$$

则$(\mathscr{H}_{x^*,y})$的解 $\hat{x}$ 存在, 且它是$(\mathscr{H}^*_{x^*,y})$的 Lagrange 乘子, 即它满足

$$f(\hat{x})+g(A\hat{x}+y)-\langle x^*, \hat{x}\rangle = \sup_{y^*}\{-f^*(-A^*y^*+x^*) -g^*(y^*)+\langle y^*, y\rangle\}$$

ii) 如果

$$y\in \operatorname{dom} g-A\operatorname{dom} f, \tag{4.6.10}$$

$$x^*\in \operatorname{int}(\operatorname{dom} f^*+A^*\operatorname{dom} g^*), \tag{4.6.11}$$

则$(\mathscr{H}^*_{x^*,y})$的解 $\hat{y}^*$ 存在, 且它是$(\mathscr{H}_{x^*,y})$的 Lagrange 乘子, 即它满足

$$-f^*(-A^*\hat{y}^*+x^*)-g^*(\hat{y}^*)+\langle y^*, y\rangle = \inf_{x}\{f(x) +g(Ax+y)-\langle x^*, x\rangle\}.$$

iii) $\hat{x}$ 是$(\mathscr{H}_{x^*,y})$的解(或 $(\mathscr{H}^*_{x^*,y})$ 的 Lrgrange 乘子)、$\hat{y}^*$ 是$(\mathscr{H}_{x^*,y})$的 Lagrange 乘子(或$(\mathscr{H}^*_{x^*,y})$的解)的充要条件为:

$$A\hat{x}+y\in \partial g^*(\hat{y}^*), \tag{4.6.12}$$

$$-A^*\hat{y}^*+x^*\in \partial f(\hat{x}). \tag{4.6.13}$$

iv) 当$(\mathscr{H}_{x^*,y})$的 Lagrange 乘子 $\hat{y}^*$ 存在时,$\hat{x}$是$(\mathscr{H}_{x^*,y})$的解等价于下列两个条件中的任何一个:

$$\hat{x}\in\partial U_y^*(x^*),\ U_y^*(x^*)=\inf_{y^*}\{f^*(-A^*y^*+x^*)+g(y^*)-\langle y^*,\ y\rangle\};\quad(4.6.14)$$

$$x^*\in\partial f(\hat{x})+A^*\partial g(A\hat{x}+y).\quad(4.6.15)$$

v) 当$(\mathscr{H}_{x^*,y}^*)$的 Lagrange 乘子 $\hat{x}$ 存在时,$\hat{y}^*$ 是$(\mathscr{H}_{x^*,y})$的解等价于下列两个条件中的任何一个:

$$\hat{y}^*\in\partial V_{x^*}(y),\ V_{x^*}(y)=\inf_{x}\{f(x)+g(Ax+y)-\langle x^*,\ x\rangle\},\quad(4.6.16)$$

$$y\in\partial g^*(\hat{y}^*)-A\partial f^*(-A^*\hat{y}^*+x^*).\quad(4.6.17)$$ ∎

定理的 i)、ii)、iii) 都可由定理 4.5.6 等得到. 其中条件(4.6.9)、(4.6.10)还可代替为对应的值有限.(4.6.14)和(4.6.16)即为(4.6.13)和(4.6.12)的对偶式

$$\hat{x}\in\partial f^*(-A^*\hat{y}^*+x^*)=\partial U_y^*(x^*),\quad(4.6.18)$$

$$\hat{y}^*\in\partial g(A\hat{x}+y)=\partial V_{x^*}(y).\quad(4.6.19)$$

(4.6.15)和(4.6.17)则可由(4.6.13)、(4.6.19) 以及 (4.6.12)、(4.6.19)得到.

这一结果当 X、Y 为自反 Banach 空间,A: $X\to Y$ 为连续线性映射时仍成立. 下列结果也一样:

定理 4.6.4 设 X、Y 为有限维空间. f: $X\to\mathbf{R}\cup\{+\infty\}$, g: $Y\to\mathbf{R}\cup\{+\infty\}$ 分别为 X、Y 上的下半连续真凸函数. A: $X\to Y$ 为线性映射. 如果

$$0\in\operatorname{int}(\operatorname{dom}g-A\operatorname{dom}f),\quad(4.6.20)$$

那么

$$\partial(f+g\circ A)(x)=\partial f(x)+A^*\partial g(Ax),\ \forall x\in X.\quad(4.6.21)$$

特别是当 $X=Y$, A 为恒等映射时,如果

$$0\in\operatorname{int}(\operatorname{dom}g-\operatorname{dom}f),$$

那么

$$\partial(f+g)(x)=\partial f(x)+\partial g(x),\ \forall x\in X;$$

当 $f\equiv 0$ 时，如果

$$0\in \operatorname{int}(\operatorname{dom} g - A(X)),$$

那么

$$\partial(g\circ A)(x)=A^*\partial g(Ax),\ \forall x\in X. \blacksquare$$

事实上，$x^*\in\partial(f+g\circ A)(\hat{x})$ 等价于 $\hat{x}$ 是问题

$$f(x)+g(Ax)-\langle x^*, x\rangle\to\min,\ x\in X$$

的解，而(4.6.20)又保证它 Lagrange 乘子存在，故由(4.6.15)，它又等价于 $x^*\in\partial f(\hat{x})+A^*\partial g(A\hat{x})$. 因此(4.6.20)成立.

定理 4.6.4 可看作 Moreau–Rockafellar 定理 4.2.3 在一定条件下的推广.

下一节介绍 Fenchel 问题的一些应用. 关系式 (4.6.12)～(4.6.19)常常可用来简化问题的计算.

§ 4.7 凸分析在其他数学分支中的应用

作为本书的最后一节，介绍一些凸分析在其他数学分支中的应用，即把本书中讨论的在一般框架中得到的结果化为在其他分支中的具体结果. 为了不过多地涉及各数学分支的预备知识，下面都只举一些尽可能简单的例子，目的仅在于指出凸分析的各种应用的可能性.

1° 变分学

变分学研究积分泛函的极值问题. 通常被积函数都要求一定的光滑性. 利用凸分析的工具，可以把原有的结果推广到凸的情形.

令

$$X=C_0^1([0, T]; \mathbf{R}^n)=\{x(t)=(x^1(t), \cdots, x^n(t))\,|\,x(t)$$
在 $[0, T]$ 上连续可微，且 $x(0)=x(T)=0\}$,

$$\|x\|_{C_0^1}=\max_{t\in[0,T]}\{\|x(t)\|, \|\dot{x}(t)\|\}$$ （这里 $\|\cdot\|$ 是 $\mathbf{R}^n$ 上的范数），

$$Y=C([0, T]; \mathbf{R}^n)=\{y(t)=(y^1(t), \cdots, y^n(t))\,|\,y(t)$$
在 $[0, T]$ 上连续$\}$,

$$\|y\|_C=\max_{t\in[0,T]}\|y(t)\|.$$

$$A: X\to Y \quad 为 \quad Ax=\dot{x}=\frac{dx}{dt}(\cdot).$$

$$F(x, y)=\int_0^T L(t, x(t), y(t))dt,$$

这里 $L: \mathbf{R}\times\mathbf{R}^n\times\mathbf{R}^n\to\mathbf{R}\cup\{+\infty\}$ 对于 x、y 为凸函数，且假定上述积分总存在(我们不深究这点成立所需要的条件，例如可假定 L 有限连续，)．于是问题

$(\mathscr{F})$

$$F(x, Ax)\to\min, \ x\in X,$$

即

$(\mathscr{F})$

$$\int_0^T L(t, x(t), \dot{x}(t))dt\to\min, \ x\in C_0^1([0, T], \mathbf{R}^n).$$

这样形式的问题起源于力学，这里涉及一个有 n 个自由度的系统，t 为时间，$x=(x^1, \cdots, x^n)$ 代表这个系统的位置，$\dot{x}=(\dot{x}^1, \cdots, \dot{x}^n)$ 代表这个系统的速度，L 称为系统的 Lagrange 函数(不要与数学规划论中的 Lagrange 函数混淆!)，它的物理意义是系统的动能与位能的差．积分称为系统的作用量，系统的轨迹 $x(t)$ 可通过上述问题$(\mathscr{F})$的求解而得到．在力学上称为最小作用量原理．

问题$(\mathscr{F})$的 Hamilton 函数在这种情形为

$$\begin{aligned}H(x, y^*)&=\sup_{y\in Y}\{\langle y^*, y\rangle-F(x, y)\}\\&=\sup_{y\in Y}\int_0^T\{\langle y^*(t), y(t)\rangle-L(t, x(t), y(t))\}dt\\&=\int_0^T\widetilde{H}(t, x(t), y^*(t))dt\end{aligned}$$

这里 $y^*\in Y^*=(C([0, T]; \mathbf{R}^n))^*$，可以证明它一定是某种测度，从而它的作用可表示为积分形式．$\widetilde{H}(t, x, y^*)$ 在变分学和力学中常称为系统的 Hamilton 函数，其物理意义为系统的总能量，其中 y^* 这时的物理意义为系统的动量．但这里与$(\mathscr{F})$的 Hamilton 函数是两个概念．

根据前面的讨论，在一定条件下有形为(4.5.4)、(4.5.17)的 Euler–Lagrange 方程和 Hamilton 方程组．为看出它们所对应的经典形式，假定系统的 Lagrange 函数 $L=L(t, x, y)$ 对于 $2n+1$ 个变量可微，则不难验证

$$\begin{aligned}\partial F(x, y) &= \{\nabla F(x, y)\}\\ &= \{(L_x(t, x, y), L_y(t, x, y))\}\subset X^*\times Y^*,\end{aligned}$$

这里

$$L_x(t, x, y)=\left(\frac{\partial L}{\partial x^1}(t, x, y), \cdots, \frac{\partial L}{\partial x^n}(t, x, y)\right),$$

$$L_y(t, x, y)=\left(\frac{\partial L}{\partial y^1}(t, x, y), \cdots, \frac{\partial L}{\partial y^n}(t, x, y)\right).$$

另一方面，由形式的分部积分可得

$$\begin{aligned}\langle y^*, Ax\rangle &= \int_0^T\left\langle y^*(t), \frac{d}{dt}x(t)\right\rangle dt\\ &=\langle y^*(t), x(t)\rangle\Big|_0^T-\int_0^T\left\langle\frac{d}{dt}y^*(t), x(t)\right\rangle dt\\ &=\int_0^T\left\langle -\frac{d}{dt}y^*(t), x(t)\right\rangle dt,\end{aligned}$$

故 A^* 可形式地理解为 $-\frac{d}{dt}$（当 y^* 可导时，$A^*y^*=-\frac{d}{dt}y^*$ 按常义理解，在一般情况下，这是一种广义导数）．这样一来，(4.5.4)就意味着存在 $\hat{y}^*\in Y^*$，满足

$$\frac{d}{dt}\hat{y}^*(t)=L_x(t, \hat{x}(t), A\hat{x}(t))=L_x(t, \hat{x}(t), \dot{\hat{x}}(t)),$$

$$\hat{y}^*(t)=L_y(t, \hat{x}(t), A\hat{x}(t))=L_{\dot{x}}(t, \hat{x}(t), \dot{\hat{x}}(t)).$$

消去 $\hat{y}^*(t)$，即得到变分学中熟知的 Euler–Lagrange 方程：

$$-\frac{d}{dt}L_{\dot{x}}(t, x(t), \dot{\hat{x}}(t))+L_x(t, \hat{x}(t), \dot{\hat{x}}(t))=0.$$

这也就说明了(4.5.4)的名称的来历．

类似地，在 $\widetilde{H}$ 可微时，也有

$$\partial_{y^*}H(x, y^*)=\{\widetilde{H}_{y^*}(t, x, y^*)\},$$

$$\partial_x(-H(x, y^*))=\{-\widetilde{H}_x(t, x, y^*)\}.$$

于是(4.5.17)变为

$$\begin{cases} \dfrac{d}{dt}\hat{x}(t)=\widetilde{H}_{y^*}(t,\ \hat{x}(t),\ \hat{y}^*(t)), \\ \dfrac{d}{dt}\hat{y}^*(t)=-\widetilde{H}_x(t,\ \hat{x}(t),\ \hat{y}^*(t)). \end{cases}$$

这是变分学中熟知的 Hamilton 方程组，从而说明了(4.5.17)的名称的来源。

我们可以设想前面关于对偶性、摄动问题等的讨论都可以在变分学中得到种种应用。

2° 最优控制

上节的变分学问题也可记作下列形式：

$$\begin{cases} \int_0^T L(t,\ x(t),\ y(t))dt\to\min, \\ \dot{x}(t)=y(t),\ x\in C_0^1([0,\ T];\ \mathbf{R}^n). \end{cases}$$

比它更一般的问题为

$(\mathscr{P})$

$$\begin{cases} J(x,\ u)=\int_0^T \varphi(t,\ x(t),\ u(t))dt\to\min; \\ \dot{x}(t)=F(t)x(t)+G(t)u(t); \\ x\in C_0^1([0,\ T];\ \mathbf{R}^n),\ u\in C([0,\ T];\ \mathbf{R}^m). \end{cases}$$

这里 $F(t)$: $\mathbf{R}^n\to\mathbf{R}^n$, $G(t)$: $\mathbf{R}^m\to\mathbf{R}^n$ 都是线性映射，作为 t 的函数都是矩阵值函数，假定它们都是连续的。

问题$(\mathscr{P})$是一个控制论问题，或者更确切地说，是一个线性系统的最优控制问题，这里线性系统指微分方程是线性的，$x(t)$称为系统的状态；$u(t)$称为系统的控制。把它抽象化，令

$$Z=C_0^1([0,\ T];\ \mathbf{R}^n)\times C([0,\ T];\ \mathbf{R}^m)$$

$$J: Z\to\mathbf{R}$$

定义为

$$J(z)=J(x,\ u)=\int_0^T \varphi(t,\ x(t),\ u(t))dt,$$

$$A: Z\to C([0,\ T];\ \mathbf{R}^n)$$

定义为

$$Az = A(x, u) = \left(\frac{d}{dt} - F\right)x - Gu.$$

于是问题($\mathscr{P}$)的抽象形式为

($\mathscr{P}$)

$$\begin{cases} J(z) \to \min, \\ Az = 0, \ z \in Z. \end{cases}$$

这是一个函数空间上的有线性等式约束的规划问题. 当J关于z是凸函数时, 我们就可应用前面的结果. 为了更直接地套用前面的结果, 我们把($\mathscr{P}$)写成

($\mathscr{P}$)

$$J(z) + \delta_0(Az) \to \min, \ z \in Z.$$

于是($\mathscr{P}$)就成了 Fenchel 问题. 注意到 $\delta_0^* \equiv 0$, $A^*: (C([0, T]; \mathbf{R}^n))^* \to Z^* = (C_0^1([0, T]; \mathbf{R}^n))^* \times (C([0, T]; \mathbf{R}^m))^*$为

$$A^* p = \left(\left(-\frac{d}{dt} - F^*\right)p, \ -G^* p\right).$$

这里 $p \in (C([0, T]; \mathbf{R}^n))^*$, $F^* = F^*(t)$, $G^* = G^*(t)$分别为$F(t)$、$G(t)$的转置, 可以验证

$$\begin{aligned} \partial J(z) &= \partial J(x, u) \\ &= (\partial_x \varphi(t, x(t), u(t)), \ \partial_u \varphi(t, x(t), u(t))), \end{aligned}$$

由(4.6.12)、(4.6.13) 可得 $\hat{z} = (\hat{x}, \hat{u}) \in Z = C_0^1([0, T]; \mathbf{R}^n) \times C([0, T]); \mathbf{R}^m)$为($\mathscr{P}$)的解, $\hat{p} \in (C([0, T]; \mathbf{R}^n))^*$ 为($\mathscr{P}$)的 Lagrange 乘子的充要条件为

$$A\hat{z} \in \partial \delta_0^*(\hat{p}) = \{0\}, \ -A^* \hat{p} \in \partial J(\hat{z}),$$

即

$$(T) \begin{cases} \dot{\hat{x}}(t) = F(t)\hat{x}(t) + G(t)u(t), \ \hat{x}(0) = \hat{x}(T) = 0; \\ \dot{\hat{p}}(t) \in -F^*(t)\hat{p}(t) + \partial_x \varphi(t, \hat{x}(t), \hat{u}(t)), \\ G^*(t)\hat{p}(t) \in \partial_u \varphi(t, \hat{x}(t), \hat{u}(t)). \end{cases}$$

条件(T)称为最优控制问题($\mathscr{P}$)的最优化条件, $\hat{p}$ 称为系统的协状态. 当φ关于x、u可微时, 它们变为

$$\begin{aligned} &\dot{\hat{p}}(t) = -F^*(t)\hat{x}(t) + \varphi_x(t, \hat{x}(t), \hat{u}(t)), \\ &G^*(t)\hat{u}(t) = \varphi_u(t, \hat{x}(t), \hat{u}(t)). \end{aligned}$$

尤其是当

$$J(x, u)=\frac{1}{2}\int_0^T\langle Q(t)x(t), x(t)\rangle dt$$
$$+\frac{1}{2}\int_0^T\langle R(t)u(t), u(t)\rangle dt$$

时，这里 $Q(t)$: $\mathbf{R}^n\to\mathbf{R}^{n*}$ 为非负定对称的，$\mathbf{R}(t)$: $\mathbf{R}^m\to\mathbf{R}^{m*}$ 为正定对称的．于是有

$$\varphi_x(t, \hat{x}(t), \hat{u}(t))=Q(t)\hat{x}(t),$$
$$\varphi_u(t, \hat{x}(t), \hat{u}(t))=R(t)\hat{u}(t).$$

从而 (T) 将变为

$$\begin{cases}\dot{\hat{x}}(t)=F(t)\hat{x}(t)+G(t)\hat{u}(t),\ x(0)=x(T)=0,\\ -\dot{\hat{p}}(t)=F^*(t)\hat{p}(t)-Q(t)\hat{u}(t),\\ G^*(t)\hat{p}(t)=R(t)\hat{u}(t).\end{cases}$$

由最后一式，

$$\hat{u}(t)=R^{-1}(t)G^*(t)\hat{p}(t),$$

从而又有

$$\begin{cases}\dot{\hat{x}}(t)=F(t)\hat{x}(t)+G(t)R^{-1}(t)G^*(t)\hat{p}(t),\\ \quad x(0)=X(T)=0,\\ -\dot{\hat{p}}(t)=F^*(t)\hat{p}(t)-Q(t)\hat{x}(t).\end{cases}$$

这些都是最优控制理论中熟知的结果．

此外，引进下列形式的 Hamilton 函数（它的含义与以前不同）H: $C_0^1([0, T], \mathbf{R}^n)\times C([0, T]; \mathbf{R}^m)^*\to\mathbf{R}\cup\{+\infty\}$，

$$H(x, u^*)=\sup_u\{\langle u^*, u\rangle-J(x, u)\}.$$

那么 $-A^*\hat{p}=\left(\left(\frac{d}{dt}+F^*\right)\hat{p}, G^*\hat{p}\right)\in\partial J(\hat{x}, \hat{u})$ 将转化为

$$\begin{cases}\hat{u}\in\partial_{u^*}H(\hat{x}, G^*\hat{p}),\\ \dot{\hat{p}}+F^*\hat{p}\in\partial_x(-H(\hat{x}, G^*\hat{p}))\end{cases}$$

这里第一个条件意味着

$$\langle G^*\hat{p}, \hat{u}\rangle-J(\hat{x}, \hat{u})=\langle\hat{p}, G\hat{u}\rangle-J(\hat{x}, \hat{u})$$
$$=H(\hat{x}, G^*\hat{p})=\sup_u\{\langle G^*\hat{p}, u\rangle-J(\hat{x}, u)\}$$

$$=\sup_u\{\langle\hat{p},\ Gu\rangle-J(\hat{x},\ u)\},$$

即 $u=\hat{u}$ 使

$$\langle\hat{p},Gu\rangle-J(\hat{x},\ u)=\int_0^T\langle\hat{x}(t),\ G(t)u(t)\rangle dt-\int_0^T\varphi(t,\ \hat{x}(t),\ u(t))dt$$

达到最大. 它相当于最优控制理论中的积分形式的 Понтрягин 最大值原理. 局部形式的 Понтрягин 最大值原理可由最优化条件 (T) 的第三个条件 $G^*(t)\hat{p}(t)\in\partial_u\varphi(t,\ \hat{x}(t),\ \hat{u}(t))$ 得到. 这时引入局部 Hamilton 函数

$$\widetilde{H}(t,\ x(t),\ u^*(t))=\sup_{u\in R^m}\{\langle u^*(t),\ u\rangle-\varphi(t,\ x(t),\ u)\},$$

则由上述条件可得

$$\hat{u}(t)\in\partial_{u^*}\widetilde{H}(t,\ \hat{x}(t),\ G^*(t)\hat{p}(t)),$$

即

$$\langle\hat{p}(t),\ G(t)\hat{u}(t)\rangle-\varphi(t,\ \hat{x}(t),\ \hat{u}(t))=\sup_u\{\langle\hat{p}(t),\ G(t)u\rangle-\varphi(t,\ \hat{x}(t),\ u)\}.$$

这就是局部形式的 Понтрягин 最大值原理. 需要注意的是这里定义的 Hamilton 函数与最优控制理论中提到的 Hamilton 函数是有所不同的.

在以上两小节中, 为了简单起见, 都假定 $x(0)=x(T)=0$. 在 $x(t)$ 不取零端点值时的其他变分问题和最优控制问题也可类似地应用凸分析的工具.

3° 偏微分方程

如果把前两小节中的常微分算子换为更复杂的函数空间中的偏微分算子, 就可把凸分析工具应用到偏微分方程问题上.

考虑下列简单的 Dirichlet 问题:

$$(D)\qquad\begin{cases}-\triangle u(x)=f(x), & \text{在 } \Omega \text{ 中};\\ u(x)=0, & \text{在 } \partial\Omega \text{ 上}.\end{cases}$$

这里 Ω 为 $\mathbf{R}^n$ 中的有光滑边界 $\partial\Omega$ 的有界区域,

$$\triangle=\sum_{i=1}^n\left(\frac{\partial}{\partial x^i}\right)^2$$

为 n 维 Laplace 算子. 令

$$L^2(\Omega)=\{\Omega \text{ 上的平方可积函数全体}\},$$

$$\|v\|_{L^2(\Omega)}=\left(\int_\Omega |v|^2 dx\right)^{1/2};$$

$$H_0^1(\Omega)=\left\{v\in L^2(\Omega)\ \middle|\ \frac{\partial v}{\partial x^i}\in L^2(\Omega),\right.$$

$$\left. i=1,\ \cdots,\ n,\ v\Big|_{\partial\Omega}=0\right\};$$

$$\|v\|_{H_0^1(\Omega)}=\left(\|v\|^2_{L^2(\Omega)}+\sum_{i=1}^n\left\|\frac{\partial v}{\partial x^i}\right\|^2_{L^2(\Omega)}\right)^{1/2};$$

$$H^{-1}(\Omega)=(H_0^1(\Omega))^*.$$

它们都是所谓 Соболев 空间. 设 $f\in H^{-1}(\Omega)$, 定义 $F: H_0^1(\Omega)\to\mathbf{R}$ 为

$$F(v)=\langle -f,\ v\rangle=-\int_\Omega f v dx,\quad \forall v\in H_0^1(\Omega).$$

$A: H_0^1(\Omega)\to(L^2(\Omega))^n$ 为线性映射:

$$Av=\operatorname{grad} v=\left(\frac{\partial v}{\partial x^1},\ \cdots,\ \frac{\partial v}{\partial x^n}\right),\ \forall v\in H_0^1(\Omega).$$

$G: (L^2(\Omega))^n\to\mathbf{R}$ 为

$$G(q)=\frac{1}{2}\sum_{i=1}^n\int_\Omega q_i^2 dx=\|q\|^2_{(L^2(\Omega))^n},$$

$$\forall q=(q_1,\ \cdots,\ q_n)\in(L^2(\Omega))^n.$$

于是问题 (D) 可归结为下列 Fenchel 问题:

$$(\mathscr{D})\qquad \begin{cases} F(v)+G(Av)\to\min,\\ v\in H_0^1(\Omega). \end{cases}$$

由 Green 公式容易验证, $A^*: (L^2(\Omega))^{n*}=(L^2(\Omega))^n\to H^{-1}(\Omega)$ 为

$$A^*q=-\operatorname{div} q=-\sum_{i=1}^n\frac{\partial q_i}{\partial x^i},$$

$$\forall q=(q_1,\ \cdots,\ q_n)\in(L^2(\Omega))^n.$$

$F^*: H^{-1}(\Omega)\to\mathbf{R}\cup\{+\infty\}$ 为

$$F^*(\varphi)=\delta_{-f}(\varphi)=\begin{cases} 0, & \text{当 } \varphi=-f;\\ +\infty, & \text{当 } \varphi\neq -f. \end{cases}$$

$G^*: (L^2(\Omega))^n\to\mathbf{R}$ 为

$$G^*(q)=\sup_s\left\{\langle q,\ s\rangle-\frac{1}{2}|S|^2\right\}$$
$$=\frac{1}{2}\|q\|^2=G(q).$$

从而对应的 Euler-Lagrange 方程为

$$\begin{cases}\Lambda u\in\partial G(p);\\ -\Lambda^*p\in\partial F(u).\end{cases}$$

但 $\partial F(u)=\{-f\}$, $\partial G(p)=\{p\}$, 故上式即

$$\begin{cases}\mathrm{grad}u=p,\\ \mathrm{div}p=-f.\end{cases}$$

消去 p 即得

$$-\triangle u=-\mathrm{div\,grad}u=f,\quad u\in H_0^1(\Omega).$$

即 u 在某种意义下是问题 (D) 的解. 它可以通过求解问题 $(\mathscr{D})$:

$(\mathscr{D})$

$$\frac{1}{2}\int_\Omega|\mathrm{grad}v|^2dx-\int_\Omega fvdx\to\min,\quad v\in H_0^1(\Omega)$$

来得到, 也可以通过求解问题 $(\mathscr{D}^*)$:

$(\mathscr{D}^*)$

$$\begin{cases}-F^*(-\Lambda^*q)-G^*(q)\to\max\\ q\in(L^2(\Omega))^n\end{cases}$$

或

$(\mathscr{D}^*)$

$$\begin{cases}\dfrac{1}{2}\|q\|^2=\dfrac{1}{2}\displaystyle\int_\Omega\sum_{i=1}^n q_i^2dx\to\min\\ \mathrm{div}q=-f,\ q\in(L^2(\Omega))^n\end{cases}$$

来得到. 类似的对偶性可以对更复杂的可化为变分(极值)问题的偏微分方程问题而提出.

4° 逼近论

关于逼近论的典型问题之一的提法如下: 设 $g\in C([0,\ 1];\ \mathbf{R})$ 为 $[0,\ 1]$ 上的连续函数. 求 $x=(x^i)\in\mathbf{R}^n$, 使得对于

$$P_n(x,\ t)=x^1+x^2t+\cdots-x^nt^{n-1}$$

有

$$(A) \qquad \max_{t\in[0,1]}|g(t)-P_n(x,\ t)|\to\min.$$

这个问题可以如下抽象化：令

$$P=\{P_n(x,\ t)\mid x\in\mathbf{R}^n\}.$$

则把 $C([0,\ 1];\ \mathbf{R})$看作 Banach 空间($\|g\|=\max\limits_{t\in[0,1]}|g(t)|$)，$P\subset C([0,\ 1];\mathbf{R})$是 $C([0,\ 1];\mathbf{R})$的闭子空间．问题(A)这时归结为

$$(A) \qquad \begin{cases}\|g-P_n\|\to\min,\\ P_n\in P.\end{cases}$$

如果进一步要求多项式 P_n 的系数 $x=(x^i)$ 是 $\mathbf{R}^n$ 中的一个凸集 K 的元素，则令

$$P_K=\{P_n(x,\ t)\mid x\in K\},$$

它显然是 $C([0,\ 1];\mathbf{R})$中的凸集，而更一般的逼近论问题为

$$(\mathscr{A}) \qquad \begin{cases}\|g-P_n\|\to\min,\\ P_n\in P_K.\end{cases}$$

下面讨论一般的抽象逼近论问题．设 X 为赋范空间，X^*为它的拓扑对偶．$C\subset X$ 为 X 的凸集．$\hat{x}\in C$ 称为 $x\in X$ 到 C 的最优逼近，是指

$$\|x-\hat{x}\|=\min_{y\in C}\|x-y\|.$$

利用凸分析的工具，可以证明下列最优逼近的充要条件：

定理 4.7.1 $\hat{x}\in C$ 为 $x\in X$ 到凸集 C 的最优逼近当且仅当存在 $x_0^*\in X^*$，使得

i) $\|x_0^*\|=\sup\limits_{\|y\|=1}\langle x_0^*,\ y\rangle=\|\hat{x}-x\|$；

ii) $\langle x_0^*,\ z-x\rangle\geqslant\|\hat{x}-x\|^2,\ \forall z\in C.$

证明 显然，$\hat{x}\in C$ 为 x 到 C 的最优逼近等价于 $\hat{x}$ 是下列问题的解：

$$F(y)=\frac{1}{2}\|y-x\|^2+\delta_C(y)\to\min,\ y\in X.$$

从而它又等价于

$$0\in\partial F(\hat{x}), \tag{4.7.1}$$

但对于 $f(y)=\frac{1}{2}\|y-x\|^2$, 由 § 4.1 的例 2, 有

$$\partial f(y)=\{x^*\in X^*|\langle x^*, y-x\rangle=\|y-x\|^2=\|x^*\|^2\}, \quad (4.7.2)$$

而

$$\partial\delta_C(y)=\{x^*\in X^*|\langle x^*, z-y\rangle\leqslant 0, \ \forall z\in C\}. \quad (4.7.3)$$

再由 Moreau-Rockafellar 定理 4.2.3, 有

$$\partial F(\hat{x})=\partial f(\hat{x})+\partial\delta_C(\hat{x}),$$

故(4.7.1)等价于存在 $x_0^*\in\partial f(\hat{x})$, $-x_0^*\in\partial\delta_C(\hat{x})$. 由(4.7.2), 有

$$\|x_0^*\|=\|\hat{x}-x\|,$$

即 i) 成立. 由(4.7.2)和(4.7.3), 有

$$\langle -x_0^*, z-\hat{x}\rangle=\langle -x_0^*, z-x+x-\hat{x}\rangle\leqslant 0, \ \forall z\in C.$$

从而

$$\langle x_0^*, z-x\rangle\geqslant\langle x_0^*, \hat{x}-x\rangle=\|\hat{x}-x\|^2, \quad \forall z\in C,$$

即 ii) 成立. 反之, 易证 i) 与 ii) $\Rightarrow x_0^*\in\partial f(\hat{x})$, $-x_0^*\in\partial\delta_C(\hat{x})$. ∎

注意到 $C([0, 1]; \mathbf{R})$ 的对偶空间为 $[0, 1]$ 上的测度, 可得到下列推论:

推论 $\hat{P}_n$ 为问题($\mathscr{A}$)的解的充要条件是: 存在 $[0, 1]$ 上的测度 μ, 使得

i) $\int_0^1 d|\mu|(t)=\sup\limits_{\|f\|=1}\int_0^1 f(t)d\mu(t)=\max\limits_{t\in[0,1]}|\hat{P}_n(t)-g(t)|$;

ii) $\int_0^1 (P_n(t)-g(t))d\mu(t)\geqslant\max\limits_{t\in[0,1]}|\hat{P}_n(t)-g(t)|^2, \forall P_n\in P_K$. ∎

这个推论不受维数限制, 也不受多项式形式的限制. 即设 $\Omega\subset\mathbf{R}^m$ 为有界闭集, $X=C(\Omega; \mathbf{R})$ 为 Ω 上的连续函数全体, $C=P_K\subset X$ 为 X 中的任何凸集, 例如某一由多项式构成的凸集. 则仍能得到类似的推论.

5° 数理经济学

凸分析在数理经济学中的应用是非常广泛的. 前面已举过不少例子(例如, 利润函数是成本函数的共轭函数等). 在这小节中, 利用凸分析的工具来得出一些微观经济学的经典结果的推广.

我们把 $\mathbf{R}^m$ 解释成为财货空间, 即 $z=(z^1, \cdots, z^m)\in\mathbf{R}^m$ 的第

i 个分量 z^i 表示第 i 种财货(商品)的量，这里假定每一种财货都可以用实数计量，当它取负值时，表示消费量或投入量，当它取正值时，表示产出量. 因此，任何 $z\in\mathbf{R}^m$ 也可表示一种生产活动，其中取负值的分量构成投入丛，而取正值的分量构成产出丛. 这时，财货空间 $\mathbf{R}^m$ 的对偶空间 $\mathbf{R}^{m*}$ 就可解释为价格体系. 即 $p=(p_1, \cdots, p_m)\in\mathbf{R}^{m*}$ 的第 i 个分量 p_i 表示第 i 种财货的价格，这里假定每一种价格也都可以用实数计量，其中负价格可解释为有害财货的价格. 不过，我们一般只考虑非负价格. $\mathbf{R}^m$ 和 $\mathbf{R}^{m*}$ 的对偶积于是就可解释为在价格体系为 $p\in\mathbf{R}_+^{m*}$ 时，财货丛 $z\in\mathbf{R}^m$ 的总价值，即

$$\langle p,\ z\rangle=\sum_{i=1}^{m}p_iz^i$$

为各种财货的价值和. 尤其是当 z 代表一种生产活动时，$\langle p, z\rangle$ 恰好就是产出丛的总价值(收入)减去投入丛的总价值(成本)，即生产活动 z 的利润.

现在设 $V\subset\mathbf{R}^m$ 解释为某企业的生产集合，即 $z\in V$ 就表示某企业可行的生产活动. 则 V 的承托函数

$$\pi(p)=\sup_{z\in V}\langle p,\ z\rangle$$

就是在价格体系为 p 时，可能得到的最大利润. 它称为企业的利润函数. 由定理 3.3.5 的推论(或命题 2.3.2)，有

$$\overline{co}\,V=\{x\in\mathbf{R}^m\,|\,\langle p,\ x\rangle\leqslant\pi(p),\ \forall p\in\mathbf{R}^{m*}\}.$$

特别是当 V 为闭凸集时，它就完全由利润函数所确定. 这一结论有重要的理论意义和实际意义. 因为生产集合 V 是不容易实际测定的，而利润函数则较容易通过会计核算来测定. 而上述结果指出当 V 为闭凸集(这点通常总认为是成立的)时，两者是一一对应的.

现在假设 V 是闭凸集，δ_V 是它的指标函数. 于是又有

$$\pi(p)=\sigma_V(p)=\delta_V^*(p),$$

(参看 § 3.5 的例 3). 因此，对于在价格体系为 $\hat{p}$ 时，达到最大利润 $\pi(\hat{p})$ 的生产活动 $\hat{z}\in V$，即 $\hat{z}$ 满足

$$\langle\hat{p},\ \hat{z}\rangle=\pi(\hat{p})=\pi(\hat{p})+\delta_V(\hat{z}),$$

其充要条件为

$$\hat{z} \in \partial\pi(\hat{p}) \tag{4.7.4}$$

(参看定理 4.1.3).

(4.7.4) 是微观经济学中的 Hotelling 引理的推广. 原来的 Hotelling 引理是对可微函数提出的. 例如, 在 $m=2$ 时, 设 $z=(y, -x)$, 其中 y 为产出, x 为投入, $\pi(p)=\pi(p_y, p_x)$ 对 (p_y, p_x) 可微. 则 (4.7.4) 指出, 在价格为 $\hat{p}=(\hat{p}_y, \hat{p}_x)$ 时, 使生产达到最大利润的产出量 $\hat{y}$ 和投入量 $\hat{x}$ 应为

$$\hat{y}=\frac{\partial}{\partial p_y}\pi(\hat{p}_y, \hat{p}_x), \quad -\hat{x}=\frac{\partial}{\partial p_x}\pi(\hat{p}_y, \hat{p}_x).$$

即它们都是利润关于价格的变化率.

如果对产出丛 y 有固定的要求, 且设 y 的维数为 l, 投入丛 x 的维数为 $n=m-l$, $z=(y, -x)\in \mathbf{R}^l\times\mathbf{R}^n$, 则

$$V(y)=\{x\in\mathbf{R}^n \mid (y, -x)\in V\}$$

称为产出为 y 时的投入需求集. 又设这时价格体系 $p=(p_y, p_x)\in\mathbf{R}^{l*}\times\mathbf{R}^{n*}$. 则

$$C(p_x, y)=\min_{x\in V(y)}\langle p_x, x\rangle$$

称为产出为 y、投入价格为 p_x 时的成本, 作为 p_x 的函数, 成本是凹函数, 且 $-C(p_x, y)$ 就是 $V(y)$ 的承托函数. 它们之间的关系与利润函数和生产集合之间的关系一样. $C(p_x, y)$ 作为 y 的函数的共轭函数为

$$\begin{aligned} P(p_y) &= \sup_{y\in\mathbf{R}^l}\{\langle p_y, y\rangle - C(p_x, y)\} \\ &= \sup_{(y, -x)\in V}\{\langle p_y, y\rangle - \langle p_x, x\rangle\} \\ &= \pi(p_y, p_x). \end{aligned}$$

即又回到了利润函数. 最后, 对于在投入价格为 $\hat{p}_x$ 时, 使成本达到最小的投入丛 $\hat{x}\in V(y)$, 其充要条件为

$$\hat{x}\in -\partial_{p_x}(-C(p_x, y)). \tag{4.7.5}$$

(4.7.5) 是微观经济学中 Shephard 引理的推广.

第四章习题

1. 对下列 $\mathbf{R}\to\mathbf{R}\cup\{+\infty\}$ 的凸函数 f，求出 f^*，∂f，∂f^*，并验证 $(\partial f)^{-1}=\partial f^*$：

   i) $f(x)=e^x$;

   ii) $f(x)=ax^2+bx+c,\ a\geqslant 0$;

   iii) $f(x)=|x-a|+|x-b|$;

   iv) $f(x)=|x|+x^2$;

   v) $f(x)=x^4$;

   vi) $f(x)=|x|^3$;

   vii) $f(x)=\begin{cases}-\log x, & x>0,\\ +\infty, & x\leqslant 0;\end{cases}$

   viii) $f(x)=\begin{cases}x\log x, & x>0,\\ 0, & x=0,\\ +\infty, & x<0,\end{cases}$

   ix) $f(x)=\delta_{[a,b]}(x)=\begin{cases}0, & x\in[a,b],\\ +\infty, & x\notin[a,b];\end{cases}$

   x) $f(x)=\begin{cases}\dfrac{x^p}{p}, & x\geqslant 0,\\ +\infty, & x<0,\end{cases}\quad p\geqslant 1.$

2. 对下列 $\mathbf{R}^n\to\mathbf{R}\cup\{+\infty\}$ 的凸函数 f，求出 f^*，∂f，∂f^*：

   i) $f(x)=\max(|x^1|,\ \cdots,\ |x^n|)$;

   ii) $f(x)=\max(x^1,\ \cdots,\ x^n)$;

   iii) $f(x)=|x^1|+\cdots+|x^n|$;

   iv) $f(x)=\begin{cases}0, & \sum\limits_{i=1}^n|x^i|^2\leqslant 1,\\ +\infty, & \sum\limits_{i=1}^n|x^i|^2>1.\end{cases}$

3. 设 f 为拓扑线性空间 X 上的真凸函数，且具有有限连续点．试证明，其右方向导数 $\delta_+f(x;h)$ 为 (x,h) 在 $\operatorname{int}(\operatorname{dom}f)\times X$ 上的上半连续函数，即
$$\lim_{(x',h')\to(x,h)}\sup\delta_+f(x';h')\leqslant\delta_+f(x;h).$$

4. 利用上题指出，如果拓扑线性空间 X 上的真凸函数 f 在点 x_0 上有限连续，那么存在 x_0 的邻域 $U(x_0)$，使得
$$\partial f(U(x_0))=\bigcup_{x\in U(x_0)}\partial f(x)$$
为 X^* 中的 $\sigma(X^*,X)$-紧集；

5. 设 $f(x,\alpha)$, $\alpha\in A$ 为一族拓扑线性空间 X 上的真凸函数，而 A 为列紧空间. 如果存在 $x_0\in X$ 的邻域 $U(x_0)$，使得 $f(x,\alpha)$ 在 $U(x_0)\times A$ 上连续，试证明

i) 函数 $f(x)=\sup\limits_{\alpha\in A}f(x,\alpha)$ 为 X 上的真凸函数，且在 x_0 上连续.

ii)
$$\partial f(x_0)=\overline{co}\bigcup_{\alpha\in A(x_0)}\partial f(x_0,\alpha),$$

这里 $\overline{co}$ 表示 $\sigma(X^*,X)$ 闭凸包, $A(x_0)=\{\alpha\in A\,|\,f(x_0,\alpha)=f(x_0)\}$.

6. 设 f 为拓扑线性空间 X 上的真凸函数, K 为 X 中的凸集. 如果 f 在 K 中有有限连续点, 或 $\mathrm{dom}f\cap\mathrm{int}K\neq\varnothing$, 试证明, 极值问题
$$f(x)\to\min,\ x\in K$$
有解 $x_0\in K$ 的充要条件为
$$-\partial f(x_0)\cap N_K(x_0)\neq\varnothing\quad(\text{Пшеничный 条件}),$$
或
$$\exists x_0^*\in\partial f(x_0),\ \forall x\in K,\ \langle x_0^*,\ x-x_0\rangle\geqslant 0\quad(\text{变分不等式}).$$

7. 在命题 4.2.7 中, 当 $F: X\to X^*$ 为连续单值映射时, 指出
$$f(x)=\int_0^1\langle F(tx),\ x\rangle dt,$$
且它是 Gâteaux 可导的处处有限的凸函数.

8. 证明当 f 为拓扑线性空间 X 上的处处有限的连续凸函数时, ∂f 是极大单调映射(Minty 定理).

9. 设 K 为赋范空间 X 的非空凸集, $d_K(x)=\inf\limits_{y\in K}\|x-y\|$. 试指出, d_K 是 X 上的连续凸函数, 且切向锥
$$T_K(x)=\{y\in X\,|\,\delta_+d_K(x;y)\leqslant 0\},\ \forall x\in K.$$

10. 设 K_1、K_2 为分离局部凸空间 X 中的两个凸集. $K=co(K_1\cup K_2)$. 试求出 $T_K(x)$, $N_K(x)$ 与 $T_{K_1}(x)$, $T_{K_2}(x)$, $N_{K_1}(x)$, $N_{K_2}(x)$ 的关系式.

11. 试把定理 4.3.5 推广到 g 仅作为 $\overline{\mathrm{aff}(\mathrm{dom}g)}$ 上的函数在 x 上连续的情形.

12. 设 g 为拓扑线性空间 $X\times Y$ 上的真凸函数, $f(x)=\inf\limits_{y\in Y}g(x,\ y)$. 如果 $\hat{y}$ 满足 $f(x)=g(x,\ \hat{y})$, 试证明 $x^*\in\partial f(x)$ 等价于 $(x^*,\ 0)\in\partial g(x,\ \hat{y})$ (参看第三章习题 25).

13. 试把定理 4.4.1 和 4.4.2 中的 $\mathbf{R}^p$, $\mathbf{R}^q$ 推广到无限维空间情形.

14. 试对线性规划问题更具体地提出命题 4.4.3、4.4.4.

15. 设 f 为分离局部凸空间 X 上的真凸函数, $K\subset X$ 为非空闭凸集. 试指出, 极值问题

$$\begin{cases} f(x) \to \min \\ x \in K \end{cases}$$

的对偶问题为

$$\begin{cases} -f^*(x^*) \to \max \\ -x^* \in K^-, \end{cases}$$

这里 K^- 为 K 的负对偶锥。

16．试对下列极值问题利用 Fenchel 问题的结果(定理4.6.3)：

$$\begin{cases} f(x) \to \min \\ x \in \mathbf{R}^n_+. \end{cases}$$

17．设 $X = C([0, 1]; \mathbf{R})$，$K = \{x \in X \mid \forall t \in [0, 1], |x(t)| \leqslant 1\}$，$a \in X$，$g$ 为 $\mathbf{R}$ 上的下半连续凸函数．试讨论下列极值问题：

$$\begin{cases} \int_0^1 x(t)dt + g\left(\int_0^1 a(t)x(t)dt\right) \to \min, \\ x \in K. \end{cases}$$

18．试讨论下列极值问题：

$$\begin{cases} f(x) \to \min \\ Ax \in K \end{cases}$$

这里 $x \in \mathbf{R}^n$，$A: \mathbf{R}^n \to \mathbf{R}^m$ 为线性映射，$K \subset \mathbf{R}^m$ 为闭凸集。

19．试讨论下列变分问题：

$$\begin{cases} \int_\Omega |\mathrm{grad}\, u|^2 dx - \int_\Omega fu dx, \\ u \in K = \{u \in H_0^1(\Omega) \mid u(x) \geqslant 0, \text{对于几乎处处的 } x \in \Omega\}. \end{cases}$$

20．把§ 4.7的4°中的问题(A)看作 $\mathbf{R}^n$ 中的极值问题

$$\begin{cases} \max\limits_{t \in [0,1]} \{\pm g(t) \mp P_n(x, t)\} \to \min, \\ x \in \mathbf{R}^n. \end{cases}$$

并利用习题5，试求出问题有解的充要条件。

参 考 文 献

[1] Moreau J.J., Fonctionnelles convexes, Seminaire sur les équations aux dérivées partielles, Collège de France, 1966～1967.

[2] Rookafellar R. T., Convex analysis, Princeton Univ. press, Princeton, New York, 1969 (俄译本,1973).

[3] Ekeland I. et Teman R., Analyse convexe et problèmes variationnels, Dunod-Gauthier-Villars, 1974 (英译本, 1976, 俄译本, 1979).

[4] Иоффе А. Д. и Тихомиров В. М., Теория экстремальных задач, Наука, Москва, 1974(英译本, 1979).

[5] Barbu V. and Precupanu Th., Convexity and optimization in Banach spaces, Romania International Publishes, Bucuresti, 1978(罗文本,1975).

[6] Aubin J. P., Mathematical methods of economic and game theory, North Holland, Amsterdam-New York Oxford, 1979. Revised ed. 1982.

[7] Пшеничный Б. Н., Выпуклый анализ и экстремальные задачи, Наука, Москва, 1980.

[8] Кутателадзе С. С. и Рубинов А. М., Двойственность минковского и ее приложения, Наука, Новосибирск, 1976.

[9] Berge C., Espaces topologiques et fonctions multivoques, Dunod, Paris, 1959(英译本, 1963).

[10] Köthe G., Topologische lineare Räume, I, Springer-Verlag, Berlin-Gottingen-Heidelberg, 1960 (英译本 1962).

[11] Ekeland I., La théorie des jeux et ses applications à l'économie mathématique, Presses Univ. de France, Collection sup., 1974.

[12] Holmes R. B., Geometric functional analysis and its applications, Springer-Verlag, New York, 1975.

[13] Aubin J. P., L'analyse non linéaire et ses motivations économiques, Masson, Paris, 1984.

[14] Bourbaki N., Espaces vectoriels topologiques, vol. I, II, Hermann, Paris, 1963.

[15] Clarke F. H., Optimization and nonsmooth analysis, Wiley-Interscience, New York, 1983.

[16] Aubin J. P. and Ekeland I., Applied nonlinear analysis, Wiley-Interscience, New York, 1984.